Niveau(x) d'enseignement: ? Numéro EDUQ : 8470478
Centre d'intérêt: 570

Chabot, Marie.

La validité du test PERPE : examen critique du contenu du test, de l'objectivité des étudiants, du concept d'insatisfaction / Marie Chabot. -- Montréal : PERPE, [1970?].
94 p. ; 28 cm.

En tête du titre: Texte B-440. Document: 0570-02. Bibliogr.: p. 93-94.

Descripteur(s): *Evaluation de l'enseignant *Evaluation par l'étudiant *Satisfaction de l'étudiant *Attitude de l'enseignant *Relations maître-élèves *Questionnaire *Evaluation de tests *Test *Validité *Collégial (Niveau)

Identificateur(s) de nom: *PERPE (Test)

Le document vise à vérifier à quel point le test PERPE (Perceptions étudiantes de la relation professeur-étudiants) répond aux objectifs qui lui ont été donnés. Il comporte quatre parties: 1. l'étude de la validité de contenu permet de constater si l'ensemble des questions décrit effectivement la relation enseignant-étudiants; 2. la seconde partie porte sur la valeur des étudiants en tant qu'observateurs, c'est-à-dire sur leur objectivité; 3. l'examen de la validité de construit concerne l'interprétation des réponses immédiates en tant qu'indices d'une insatisfaction; 4. l'étude de la validité concomitante et prédictive consiste à évaluer sa capacité de prédire le comportement ou la réussite d'un enseignant à une date ultérieure.

En ce qui concerne la validité de contenu, il apparaît que la relation enseignant-étudiants est adéquatement représentée par les 61 dimensions du test et que l'opérationalité et l'unidimensionalité des dimensions favorisent la mesure. Quant aux perceptions étudiantes, elles ne sont affectées ni par la réussite scolaire personnelle ni par l'âge. La classification des scores associés à l'insatisfaction étudiante a abouti à des sous-groupes cohérents et unifiés d'enseignants et d'étudiants. Enfin, l'exploration des conditions préalables des validations prédictive et concomitante conduit à envisager celles-ci comme ultérieures.

Type de document: r

EDUQ 8470478

3800 est, rue Sherbrooke
Montreal, 406 (254-7131)

Projet **P** erceptions **E** tudiantes de la **R** elation **P** rofesseur **E** tudiants

TEXTE B-440

LA VALIDITÉ DU TEST PERPE

examen critique

- du contenu du test
- de l'objectivité des étudiants
- du concept d'insatisfaction

par
MARIE CHABOT
D.Ph. (Ps.)

C
•181
Q

0570-02

EDUQ 8470478

147

Texte B - 440

LA VALIDITÉ DU TEST PERPE

Examen critique

- du contenu du test
- de l'objectivité des étudiants
- du concept d'insatisfaction

Marie Chabot, D. Ph.(Ps.)

Document 0570 - 02

Table des matières

	Page
Introduction	1
I - La validité de contenu	5
A - La pertinence du contenu	9
B - La présentation du contenu	26
C - L'appréciation du contenu par les professeurs	33
D - Synthèse	38
II - La valeur des étudiants en tant qu'observateurs	42
A - Les étudiants comme observateurs privilégiés	43
B - L'aptitude des étudiants à répondre au questionnaire	44
C - L'objectivité des étudiants	47
III - La validité de construit	58
A - Les notions de "réalité" et d'"attentes"	61
B - Le construit "satisfaction"	63
C - Quelques hypothèses liées au concept de satisfaction	68
D - La notion de compétence	75
IV - La validité de prédiction et la validation concomitante	81
A - La validité prédictive	82
B - La validité concomitante	85
Conclusion	89
Références	93

INTRODUCTION

L'étude de la validité est la démarche selon laquelle nous pouvons vérifier à quel point un test répond aux objectifs qui lui ont été donnés. Différents types d'objectifs commanderont des types de validité différents, et rares sont les tests psychologiques qui peuvent se réclamer d'objectifs assez variés pour réaliser toutes les conditions attenantes aux diverses formes de validité.

Le nom même du test PERPE "Perceptions Étudiantes de la Relation Professeur-Étudiants", donne déjà une indication d'au moins deux de ses objectifs majeurs. Si PERPE vise à mesurer la relation professeur-étudiants, ou en d'autres termes la situation d'enseignement, celle-ci doit être bien représentée par les item ou questions contenues dans le test. L'étude de la validité de contenu, au chapitre I, nous permettra de vérifier si l'ensemble des questions décrit effectivement la relation professeur-étudiant. Cependant, la situation d'enseignement est décrite au moyen des perceptions étudiantes; nous devons donc démontrer que les étudiants, dans la mesure où ils font partie de l'instrument de mesure, sont suffisamment objectifs pour que leurs réponses puissent être prises au sérieux. Le chapitre II, sur "La valeur des étudiants en tant qu'observateurs" nous permettra d'envisager cet aspect de la validité de notre instrument.

L'un des objectifs du test consiste en outre à renseigner le professeur sur les perceptions qu'ont les étudiants à propos de la réalité du cours et à propos de leur idéal face au cours; cependant, les informations fournies au professeur dépassent en complexité les simples réponses données par les étudiants: les réponses immédiates sont interprétées comme étant des indices d'une réponse médiatisée que nous appelons "insatisfaction". Ce processus, qui n'est pas immédiatement fourni par les réponses, est donc construit par les interprètes des résultats, en fonction d'une théorie quelconque qui justifie de telles inductions. Nous verrons au chapitre III, intitulé "La validité de construit" s'il est légitime de parler d'insatisfac-

tion dans l'interprétation des résultats. Le construit ou concept de "compétence" sera également abordé dans ce chapitre.

Au chapitre IV, portant sur la validité concommitante et prédictive, nous discuterons brièvement un objectif que le test ne prétend pas réaliser actuellement: celui de prédire le comportement ou la réussite d'un professeur à une date ultérieure. Quelques indications sur la validité concommitante mettront en lumière la capacité du test à refléter effectivement les perceptions étudiantes.

Pour simplifier la récapitulation et la saisie rapide de l'essentiel du texte, les en-tête sont la plupart du temps accompagnées de "manchettes", numérotées selon le chapitre et selon leur ordre d'apparition.

Chapitre I

LA VALIDITÉ DE CONTENU

Lorsque l'usager d'un test veut savoir "comment se comporte un individu dans un répertoire de situations que le test est censé représenter", il faut pouvoir démontrer que "le test est un bon échantillonnage des situations sur lesquelles on veut porter une conclusion" (A.P.A., 1966, p. 12)[1]. Cette phrase résume la démarche de la validité du contenu. L'un des objectifs du test PERPE est de fournir aux professeurs l'évaluation que font leurs étudiants de la relation professeur-étudiant. La validité de contenu doit démontrer que cet objectif est atteint, et par conséquent que la relation professeur-étudiant est bien représentée par les dimensions qui ont été retenues dans le test.

Cette démonstration peut utiliser des cheminements divers. "Si les futurs usagers du test ou des personnes en position de responsabilité reconnaissent que le plan est adéquat et fut bien réalisé, le test a une grande validité de contenu" (Nunnaly, 1967, p. 81). Il s'agit ici d'un argument d'autorité, qui est indispensable par exemple lorsqu'il faut démontrer que les questions d'un examen sont bien représentatives de la matière vue au cours. Une autre voie possible pour exposer la validité de contenu consiste à exploiter les mesures de corrélation entre un contenu nouveau et d'autres contenus portant sur le même domaine. Cette méthode risque de simplement faire apparaître une relation entre deux contenus de valeur douteuse. "Malgré les efforts de certains pour régler tous les problèmes de mesure psychologique par une fuite dans la statistique, la validité de contenu est surtout établie par d'autres moyens... la validité de contenu repose surtout sur l'exposé de la pertinence du contenu et de la façon dont il est présenté."(Nunnaly, 1967, p. 83).

---

[1] Les citations d'un texte anglais sont traduites par l'auteur.

En fait, la voie statistique la plus usuelle pour valider un contenu, celle de la mise en corrélation d'une mesure avec des contenus semblables, est impraticable dans l'état actuel des recherches en pédagogie. Le chercheur américain Ryans a publié en 1963 une liste des variables utilisées par une variété d'auteurs, entre 1958 et 1963 seulement, pour caractériser la situation d'enseignement. Au seul chapitre des caractéristiques du professeur, on peut compter 91 variables! (Ryans, 1963, pp. 434-437). Les travaux de synthèse sont si peu avancés dans ce domaine qu'il aurait été vain de vouloir comparer nos résultats à ceux d'autres chercheurs, d'autant plus que le recours aux étudiants comme évaluateurs est rarement utilisé.

Pour étudier la validité du contenu du test PERPE, nous adopterons donc la voie suivante: nous discuterons

A - <u>La pertinence du contenu</u>, sous trois en-têtes

1. L'origine des dimensions
2. La classification des dimensions
3. La représentativité des dimensions: synthèse.

Cette approche nous permettra, croyons-nous, d'établir dans quelle mesure les dimensions du test représentent adéquatement la situation d'enseignement. En un second temps nous discuterons

B - <u>La présentation du contenu</u>, sous deux en-têtes

1. L'opérationalité des dimensions
2. L'unidimensionalité des dimensions.

Nous montrerons, lors de ces deux sous-sections, dans quelle mesure la présentation du contenu est pertinente à l'<u>observation</u> et à la <u>notation</u> des comportements professoraux. En dernier lieu nous examinerons les témoignages fournis par les professeurs en 1968 et en 1969:

C - <u>L'appréciation du contenu par les professeurs</u>

1. Les entrevues de 1968
2. L'enquête de 1969.

Nous terminons par une 4ème section intitulée:

D - <u>Synthèse générale</u>.

A - LA PERTINENCE DU CONTENU

1. L'origine des dimensions

1.1. Les dimensions servant à évaluer la relation professeur-étudiant ont été élaborées à l'origine par un groupe d'étudiants de niveau collégial.

1.2. Quatre versions du test, continuellement perfectionnées, furent mises à l'essai avant que la cinquième version, définitive cette fois, soit expérimentée à l'échelle du réseau de l'enseignement collégial.

1.3. Les précautions apportées à la construction des dimensions font en sorte que celles-ci se situent dans le cadre de référence des étudiants qui s'en serviront pour évaluer leurs professeurs.

1.4. En dehors des aspects spécifiques à chaque matière, le contenu reflète bien les aspects mesurables de l'enseignement actuel.

Le test PERPE débuta dans le cadre des travaux des étudiants d'un cours de psychologie sociale, de niveau baccalauréat. Le professeur, devenu par la suite l'auteur du test, proposa à ses étudiants de former des groupes de discussion, dont la tâche serait de clarifier "les critères selon lesquels on juge un cours, dans le but pratique d'établir une fiche d'évaluation, applicable à tout cours et à tout professeur". Cinq rencontres, tenues en décembre 1966 et janvier 1967, furent centrées autour des thèmes de discussion suivants:

la notion de structuration d'un cours;
les critères d'évaluation de la personnalité d'un professeur;

les critères d'évaluation d'un contenu de cours;
le concept de réponse du professeur aux attentes des étudiants;
la notion d'interaction.

A la suite de ces discussions, la rédaction des dimensions de la situation d'enseignement prit la forme suivante: chaque étudiant eut pour tâche d'observer ses professeurs et de noter des faits de comportement. Les faits à noter, bien entendu, devaient paraître importants aux yeux des étudiants.

La fiche de notation comportait

a) la description d'un comportement

b) une classification de ce comportement:

(C): relié au contenu
(I): relié à l'interaction professeur-étudiant
(P): relié à la personnalité du professeur

c) une qualification de ce comportement: positif ou négatif.

Les 600 fiches descriptives ainsi obtenues servirent de point de départ à l'auteur du test, aidé par une équipe de volontaires, pour la préparation de la première forme expérimentale. Une échelle en cinq points fut construite pour chaque dimension. Cette première forme du test (forme A) fut mise à l'essai quelques mois plus tard, en juillet '67, avec la collaboration de 24 classes d'une session d'été d'un collège de Montréal. L'accueil favorable des professeurs et l'ensemble des premiers résultats eurent pour effet de convaincre l'auteur que l'évaluation par les étudiants s'avérait prometteuse.

Seules des transformations mineures suivirent la forme A. Les commentaires des professeurs et des étudiants qui collaborèrent à l'expérience firent en sorte que la description de quelques dimensions fut reformulée; six échelles de réponses furent remaniées et cinq descriptions de dimensions retouchées afin de corriger quelques instances d'utilisation

anormale de l'échelle. Les dimensions portant sur la personnalité du professeur furent placées à la fin du test, afin de mieux garantir l'objectivité des étudiants.

Après l'expérimentation de la forme B, l'auteur et une équipe d'étudiants universitaires en psychologie remanièrent le contenu à la lumière des informations psychométriques obtenues avec les formes A et B. Deux formes parallèles furent alors mises sur pied, les formes C et D. Trente-trois dimensions nouvelles furent introduites, afin de porter à 61 le nombre de chacune des deux formes parallèles. Les dimensions mises à l'essai dans les formes C et D provenaient de quelques suggestions de professeurs déjà évalués et surtout d'un examen de la littérature (Ryans, 1963; Solomon et al.,1964). De plus, il arriva que deux dimensions de la forme B soient fusionnées en une seule dans les formes C/D ou, à l'inverse, qu'une dimension soit dédoublée.

En fin de compte, la forme F provient d'une sélection des meilleures dimensions des formes A/B et C/D[1]. En tout, 97 dimensions furent expérimentées, et puisqu'il en reste 61, 36 furent abandonnées en cours de route. Le tableau 1 indique la provenance des dimensions de la forme F. Ce tableau permet de voir que malgré les trois expérimentations qui suivirent la forme A, 37 des dimensions originelles, c'est-à-dire 60%, demeurent dans la forme finale, la plupart ayant subi de légères modifications.

Nous constatons maintenant, a posteriori, que ce fut un heureux procédé que de construire la première forme du test à l'aide de dimensions fournies par des étudiants; il est essentiel, en effet, qu'un questionnaire se situe dans le système de référence des personnes qui auront à y répondre: "Le questionnaire doit présenter chaque objet dans une forme qui correspond à la perception qu'en a l'homme qu'on interroge" (Cannell et Kahn, 1959, p. 403).

---

[1] La sélection et les critères d'évaluation des dimensions seront discutés dans le texte B - 320.

Tableau 1

Provenance des dimensions de la forme F

Dimensions apparues en A - B		37	(60%)
Dimensions apparues en C	11		
Dimensions apparues en D	6	24	(40%)
Dimensions apparues en C et D simultanément	7		
Total des dimensions de la forme F		61	

En faisant appel aux dimensions qui paraissaient être, aux yeux d'un groupe d'étudiants, d'importants aspects de la situation d'enseignement, PERPE utilisa une voie empirique qui garantissait, en quelque sorte, que les dimensions allaient avoir une réelle signification pour eux. Il s'agit là d'une qualité fondamentale pour PERPE, car si les professeurs allaient devenir les usagers des résultats, les étudiants de niveau collégial, par contre, allaient devenir les premiers usagers du test en tant que questionnaire.

Par ailleurs, si l'échantillonnage initial des dimensions provenait des perceptions étudiantes de la relation professeur-étudiant, celui-ci fut complété par la suite selon les suggestions de professeurs, qui exposèrent des éléments qu'eux voudraient voir évaluer. Ainsi, sauf pour quelques dimensions préparées à la lumière d'autres recherches, les dimensions de PERPE furent élaborées avec la collaboration des personnes impliquées par l'utilisation du test. Dans quelle mesure cette voie empirique a-t-elle produit un échantillonnage complet des éléments de la situation-cours? C'est ce que nous verrons maintenant.

2. La classification des dimensions

> 1.5. Le contenu du test couvre les zones suivantes: A - le cours lui-même, B - le professeur et C - l'interaction professeur-étudiant, avec un nombre presqu'égal de dimensions.
>
> 1.6. Pour analyser les limites du contenu, chacune des 6 sections du test doit être envisagée séparément.

La classification qui sera présentée ici vise à regrouper les dimensions en familles, de façon à donner au lecteur une vue plus synthétique de leur répertoire. Le tableau 2 fournit une vue d'ensemble des dimensions au moyen d'un regroupement qui fut adopté à l'automne 1969, par l'équipe du projet PERPE, pour faciliter l'analyse des dossiers des professeurs. Ce regroupement doit être considéré comme provisoire, puisqu'il repose uniquement sur des considérations logiques et non sur des analyses statistiques.

Dans la première colonne de gauche nous trouvons une présentation des dimensions en termes des composantes d'une communication:

A - Le MESSAGE, c'est-à-dire le cours lui-même
B - L'ÉMETTEUR, c'est-à-dire le professeur
C - L'INTERACTION professeur-étudiant.

Ce regroupement, qui constitue la classification la plus générale possible des dimensions, nous permet de voir qu'un aspect du cours n'est pas inclus dans le contenu du test: il s'agit du RÉCEPTEUR, c'est-à-dire des étudiants. En fait, lorsque nous disons que PERPE évalue la situation d'enseignement, nous visons cette situation telle qu'elle est vécue et perçue du point de vue des étudiants, en excluant leur perception d'eux-mêmes. Les traits qui leur sont spécifiques, tels leur collaboration, leur ponctualité, leur curiosité intellectuelle, etc... sont donc mis de côté et le contenu, dans son ensemble, est centré exclusivement sur le professeur.

Tableau 2

Modèle provisoire pour le regroupement
des 61 dimensions du test PERPE

Catégories	Sous-catégories	Sections*	Dimensions	N
(A) LE MESSAGE i.e. le cours	1. Ce qu'il est: caractéristiques liées au contenu du cours et aux techniques d'évaluation.	I	1 à 9	9
	2. Comment il est transmis: organisation des contenus et techniques pédagogiques.	II	10 à 20	11
(B) L'EMETTEUR i.e. le professeur	1. Ce qu'il est: attitudes et traits de personnalité ayant des incidences pédagogiques.	VI	53 à 61	9
	2. Comment il transmet: caractéristiques immédiates et médiates de l'exposé verbal.	III	21 à 30	10
(C) L'INTERACTION professeur-étudiants	1. Dimension cognitive: interactions liées à la tâche d'enseignement (professeur) et/ou d'apprentissage (étudiants).	IV	31 à 44	14
	2. Dimension socio-émotive: interactions liées au climat socio-émotif dans et hors de la classe.	V	45 à 52	8

* L'ordre adopté dans le test PERPE reporte à la fin les sections plus délicates à évaluer, telles les interactions quant à l'émotivité et la personnalité du professeur.

La deuxième colonne du tableau 2 fournit une classification en six sous-catégories, qui précisent un peu plus l'orientation du contenu. Nous reviendrons plus en détail sur chacune d'entre elles. Les numéros de section présentés à la 3è colonne indiquent la place des sections dans le test. Les sections plus délicates à évaluer, telles les interactions quant à l'émotivité (section V) et la personnalité du professeur (section VI) ont été reportées à la fin du test. Nous estimons que les sections dont les dimensions portent sur des éléments plus objectifs de l'enseignement servent d'entraînement à l'objectivité, lorsqu'elles précèdent des sections qui comportent des dimensions dont l'évaluation est plus subjective.

La quatrième colonne identifie les numéros des dimensions et la cinquième en donne le nombre par section. Nous pouvons constater que les sections sont d'égale importance quant au nombre des dimensions (N = 8 à 11), sauf la section IV (Interaction d'ordre cognitif), un peu plus considérable puisqu'elle contient 14 dimensions. Nous présenterons maintenant les sections une à une, en donnant dans chaque cas la liste des dimensions qu'elles contiennent.

a) Section I. Ce qu'est le cours: contenu et évaluation.

Liste des dimensions:

1. Quantité de matière
2. Contenu substantiel
3. Réponse aux attentes quant au contenu
4. Quantité de lectures obligatoires
5. Fréquence des contrôles
6. Questions fermées dans les contrôles
7. Questions ouvertes dans les contrôles
8. Compréhension dans les contrôles
9. Degré de mémorisation dans les contrôles.

A cause de la destination généralisée du test, celui-ci devant servir à évaluer toutes les matières, le choix des dimensions portant sur le contenu du cours est forcément très limité. Toute question spécifiquement applicable à une matière serait inapplicable aux autres matières et serait par conséquent rejetée par un trop grand nombre d'étudiants. Si nous utilisons l'expression "contenu et évaluation" pour caractériser cette

section, il convient de rappeler que le terme "contenu" n'est applicable qu'aux dimensions 1. Quantité de matière, 2. Contenu substantiel et 3. Réponses aux attentes quant au contenu. La dimension 4. Quantité de lectures obligatoires se rapporte de façon indirecte au contenu puisque les lectures constituent un moyen normal de transmission de la matière. En somme, les dimensions retenues pour caractériser le contenu traitent surtout de l'administration de celui-ci. D'autres dimensions ont été mises à l'essai dans les premières formes du test; il s'agit notamment de "degré de dédoublement du contenu", "difficulté des lectures", "parenté entre lectures et matière", "parenté entre travaux et matière" et "quantité globale de travail". Elles ont été rejetées de la sélection finale pour divers motifs, mais parmi ceux-ci, un très faible taux d'insatisfaction fut une caractéristique commune. Les étudiants, semble-t-il, ont peu de plaintes à formuler sur ces questions et lorsqu'ils en ont, les quelques insatisfaits ont tendance à se contredire beaucoup entre eux.

Le domaine du contrôle de l'apprentissage est assez bien représenté, par les dimensions 5 à 9, si l'on considère que le seul aspect majeur qui n'est pas envisagé est celui de la "difficulté des contrôles". Lorsque cette dimension fut mise à l'essai dans les formes A et B, elle occasionna si peu de différences entre les classes, quant à leur insatisfaction, qu'il fallut l'abandonner. La question de la longueur des contrôles n'est pas abordée non plus mais il est probable qu'elle aurait donné les mêmes résultats que celle de la difficulté des contrôles. Remarquons que les contrôles sont envisagés ici dans leur forme traditionnelle: il n'est pas question d'auto-évaluation ni d'évaluation par le groupe.

En conclusion à l'examen rapide de cette section, nous pouvons dire que les dimensions portant sur le contenu et l'évaluation des cours se situent à l'intérieur de contraintes précises:

a) les dimensions spécifiques à certains secteurs ayant été rejetées, celles qui ont été retenues sont uniquement celles qui décrivent des caractéristiques communes à toutes les matières;

b) les dimensions retenues sont nécessairement des aspects susceptibles de soulever de l'insatisfaction chez les étudiants; sinon l'évaluation n'offre pas de prise à la différentiation entre les professeurs;

c) l'évaluation dont il est question est une évaluation de type traditionnel: elle est faite par le professeur, au moyen d'un examen.

b) Section II. La transmission du cours: organisation des contenus et techniques pédagogiques.

Liste des dimensions:

10. Structuration
11. Informations sur l'organisation du contenu
12. Approche pédagogique
13. Type d'exposé
14. Pouvoir clarificateur des exemples
15. Synthèses
16. Planification du travail
17. Fréquence du travail en groupe
18. Matériel didactique
19. Stabilité de l'approche pédagogique
20. Réponses aux attentes quant à l'approche pédagogique.

Ces dimensions ont toutes rapport à la méthodologie de l'enseignement; toutefois, les dimensions 10 à 16 portent surtout sur la forme de présentation du cours: le cours est-il structuré? l'exposé est-il magistral? etc.... Tout en illustrant spécialement bien la technique du professeur lorsqu'il donne son cours, ces dimensions sont facilement transposables à tous les types de contenus, c'est-à-dire à toutes les matières.

La qualité de la présentation de la matière, notamment, est bien circonscrite par les dimensions suivantes: 10. Structuration, 11. Informations sur l'organisation du contenu, 14. Pouvoir clarificateur des exemples et 15. Synthèses. Notons qu'il s'agit là de qualités formelles plutôt simples; des qualités formelles plus complexes, telles que "niveau d'abstraction" ou "déduction et induction" ont donné lieu, en pré-expérimentation, à des taux de rejet et de contradiction trop élevés. Il est possible

que des dimensions comme "abstraction" ou "déduction et induction", soient difficilement mesurables parce qu'elles portent sur la démarche intellectuelle du professeur et qu'elles sont par conséquent plus éloignées de la perception des étudiants. Les dimensions 17. Fréquence du travail en groupe et 18. Matériel didactique permettent une évaluation des variations introduites par le professeur dans son enseignement. Toutefois, le détail de ces deux méthodologies n'est pas exploré. Nous pensons à des aspects plus spécifiques, tels les suivants: principes de constitution des groupes, grosseur des groupes, caractère substantiel des discussions, fréquence des communications entre les groupes, disponibilité du professeur une fois les groupes constitués, ou encore, caractère substantiel des apports audio-visuels, fréquence des discussions centrées sur des documents cinématographiques, etc.... Notons que des techniques encore moins fréquentes que le travail en groupe ou l'utilisation de moyens audio-visuels, tels que recherches dans le milieu ou visites et stages, ne sont pas mesurés par le contenu actuel de PERPE.

En résumé, cette seconde section de dimensions permet d'explorer la plupart des aspects formels d'une présentation magistrale de la matière, de même que la fréquence de deux techniques alternatives d'enseignement.

c) Section III. Le professeur: caractéristiques de l'exposé oral.

Liste des dimensions:

21. Syntaxe et vocabulaire
22. Clarté du langage
23. Débit
24. Régularité
25. Digressions désirables
26. Digressions indésirables
27. Capte l'attention
28. Capte l'attention par le ton
29. Capte l'attention par le contenu
30. Performance dramatique

L'exposé oral est caractérisé à lui seul par dix dimensions. Comment justifier que ce seul aspect du cours ait nécessité un si grand nombre

de dimensions? De fait, dès l'origine du test, les étudiants accordaient une grande importance à ce type de dimensions, puisque huit des dix dimensions actuelles portant sur l'exposé oral furent introduites lors de la première construction (forme A), et qu'aucune n'a présenté depuis de motifs d'élimination. C'est dire que non seulement ces dimensions ont semblé importantes aux yeux des étudiants, mais qu'elles soulèvent, d'une classe à l'autre, des taux d'insatisfaction à la fois assez élevés et assez variables pour bien discriminer entre les professeurs.

Bien que cette section du contenu soit surtout centrée sur "la parole" du professeur, son champ d'application n'est nullement limité au seul enseignement magistral. En effet, même dans une forme d'enseignement qui réduit au minimum l'exposé du professeur, ce dernier conserve un rôle prédominant et la composante verbale de sa présence demeure son principal outil de communication. Il est à prévoir, par conséquent, que même dans l'optique d'une transformation des techniques pédagogiques, une section comme celle-ci demeurerait l'une des plus validement applicables à toute forme d'enseignement.

d) Section VI. Le professeur: ce qu'il est.

Liste des dimensions:

53. Humeur
54. Humour
55. Objectivité
56. Jugement nuancé
57. Humilité
58. Justice
59. Intérêt pour les individus
60. Degré d'empathie
61. Culture générale

Ces neuf dimensions sont celles qui s'apparentent le plus à une description de la personnalité du professeur. Notons toutefois que d'une part, le test ne prétend pas mesurer la personnalité du professeur dans son ensemble et que d'autre part, les traits personnels inclus dans ces dimensions ne se rattachent nullement à une "théorie" quelconque de la personnalité. Il s'agit donc strictement de certains éléments de la personnalité, retenus à cause de leur lien avec le travail d'enseignement. Dans cette

section, peut-être plus que dans tout autre, la détermination du contenu fut nettement empirique: les suggestions des étudiants ont constitué un premier répertoire de dimensions et les critères de sélection ont ensuite conduit au rejet de quelques-unes. C'est ainsi que 20 dimensions portant sur des aspects personnels du professeur furent expérimentées; neuf dimensions sur 20 seulement furent conservées, parce que dans l'ensemble les étudiants expriment peu d'insatisfaction quant à la personnalité de leur professeur, et que ce "contentement" est stable d'une classe à l'autre. Le contenu de cette partie du test ne comporte aucune dimension liée à des traits physiques ou extérieurs du professeur, parce que justement la mesure de ces traits ne fournit aucun renseignement intéressant. Parmi celles qui furent éliminées, donnons en exemple les dimensions suivantes: "ponctualité", "mobilité", "assurance", "propreté", "manies et tics", caractéristiques physiques" et "maintien". Des dimensions portant sur la compétence ou sur l'éthique professionnelle n'apparaissent pas au contenu pour les mêmes raisons; nous faisons allusion ici à des dimensions rejetées telles que "engagement apparent face à la profession", "conscience professionnelle", "compromission" ou "compétence académique".

Les dimensions effectivement conservées peuvent se regrouper de la facon suivante:

1. Qualités intellectuelles:	54. Humour 55. Objectivité 56. Jugement nuancé 61. Culture générale
2. Qualités interpersonnelles:	58. Justice 59. Intérêt pour les individus 60. Degré d'empathie
3. Qualités personnelles:	53. Humeur 57. Humilité

Le contenu du test fait donc une maigre part aux qualités réellement personnelles des professeurs et donne priorité aux qualités intellectuelles et interpersonnelles. Par ailleurs l'évaluation des qualités professionnelles aussi bien que celle des traits physiques ou extérieurs est absente.

e) Section IV. L'interaction cognitive.

Liste des dimensions:

31. Intellectuellement stimulant
32. Aptitude à clarifier
33. Compréhension des questions
34. Concrétisation (encourage la...)
35. Attentif aux aptitudes individuelles
36. Patience pédagogique
37. Attitude valorisatrice
38. Attitude démocratique
39. Fréquence de la critique
40. Nature de la critique
41. Fréquence des questions par le professeur
42. Fréquence des questions par les étudiants
43. Responsabilité face à la fréquence réduite des questions
44. Fréquence des échanges d'opinions.

Nous venons de voir plus haut que la section IV (Le professeur: ce qu'il est) offre un éventail assez restreint de traits de personnalité, et que trois de ces traits peuvent être envisagés comme des qualités interpersonnelles (58. Justice; 59. Intérêt pour les individus; 60. Degré d'empathie). Nous faisions remarquer en outre que les dimensions retenues dans cette section ne faisaient nullement état d'une théorie de la personnalité. Dans la présente section, par contre, nous retrouvons plusieurs qualités simultanément interpersonnelles et pédagogiques. Ces qualités sont considérées comme devant avoir un effet positif sur l'apprentissage de l'étudiant.

Exemples:

31. Intellectuellement stimulant
32. Aptitude à clarifier
34. Concrétisation (encourage la...)
35. Attentif aux aptitudes individuelles
36. Patience pédagogique
37. Attitude valorisatrice
38. Attitude démocratique
40. Nature de la critique (constructive ou non)

Il est clair que ce groupe de dimensions participe d'une option théorique quant à la nature d'une "saine" pédagogie: celle-ci est vue comme devant être centrée sur les étudiants plutôt que sur le professeur, et non seulement sur les étudiants comme groupe, mais en tant qu'individus également.

Les 8 dimensions que nous venons d'énumérer circonscrivent surtout des "attitudes" pédagogiques; elles trouvent leur complément dans un quatuor de dimensions évaluant des "comportements" plus directement observables:

33. Compréhension des questions
39. Fréquence de la critique
41. Fréquence des questions par le professeur
44. Fréquence des échanges d'opinions.

L'on peut se demander si l'objectivité du test n'aurait pas bénéficié d'une concentration plus marquée de dimensions de "comportement". En fait, les premières formes du test nous ont permis de constater que plusieurs aspects "objectifs" de l'interaction cognitive (par exemple: "l'évaluation des questions des étudiants", "rigueur d'expression demandée", "contrôle de la durée des questions", "parenté entre échanges et contenu") occasionnent plusieurs rejets, suscitent peu d'insatisfaction et offrent en général peu de prise à la mesure. Ainsi, l'actuel contenu est centré sur des attitudes plutôt que sur des comportements autant par nécessité psychométrique que par tendance théorique.

Une des dimensions de cette section échappe à la formulation générale du contenu voulant que les dimensions ne portent pas sur les étudiants; il s'agit de la dimension 42. Fréquence des questions par les étudiants. Dans la section V (Interaction émotive) nous retrouverons une autre dimension évaluant l'étudiant: 46. Agressivité verbale par les étudiants. De fait, ces deux dimensions constituent un échantillon si limité du comportement des étudiants, qu'il conviendrait soit de les éliminer, soit de les mettre dans une section à part, en y adjoignant plusieurs autres dimensions. L'on peut imaginer, dans l'optique du développement et du perfectionnement ultérieur du contenu, que la perception qu'ont les étudiants d'eux-mêmes constituerait un volet complémentaire du test, venant enrichir les informations sur leur perception du cours et du professeur. Remarquons toutefois, à la décharge de cette inconsistance dans le contenu du test, que ces deux comportements étudiants sont liés à l'action du professeur, qui en provoque et en contrôle les manifestations.

Ajoutons pour conclure que si l'actuel contenu est centré sur des aspects d'interaction délicats et difficiles à connaître, son champ d'application dépasse pour cette raison l'enseignement magistral, et peut englober toute forme de cours.

f) Section V. Interaction socio-émotive.

Liste des dimensions:

45. Climat interpersonnel
46. Agressivité verbale des étudiants
47. Forme de discipline
48. Ponctualité attendue
49. Contacts hors cours (terrain académique)
50. Contacts hors cours (terrain non-académique)
51. Initiative des contacts hors cours
52. Nature des échanges hors cours.

Ces dimensions regroupées sous l'en-tête "Interaction socio-émotive" constituent un sous-groupe de dimensions plus éloignées de l'apprentissage que les autres. D'ailleurs, les quatre dernières de la liste (dimensions 49 à 52) traitent de ce qui se passe en dehors du cours. L'accent placé sur l'enseignement magistral dans les sections I (contenu et évaluation), II (Méthodologie du cours) et III (Caractéristiques de l'exposé oral) est donc partiellement mis de côté ici, puisque la relation professeur-étudiant est envisagée comme débordant les heures de cours.

Nous trouvons dans cette section les deux seules dimensions du test qui rejoignent des activités disciplinaires du professeur (47. Forme de discipline et 48. Ponctualité attendue). Une seule autre dimension de ce type fut mise à l'essai en pré-expérimentation, et ni les étudiants ni les professeurs ne se sont plaints d'une pénurie de ce côté. Le rôle disciplinaire du professeur apparaît donc comme n'étant plus d'actualité, au niveau collégial du moins, et deux dimensions suffisent à le mesurer.

3. La représentativité des dimensions: synthèse.

Que pouvons-nous dire, en conclusion, sur le répertoire des dimensions contenues dans le test PERPE? Il ressort du rapide coup d'oeil que nous venons de jeter sur les dimensions, que PERPE peut être utilisé validement pour évaluer plusieurs aspects de l'enseignement contemporain, mais qu'il comporte néanmoins des limitations.

Ce que PERPE peut évaluer validement:

1. PERPE évalue le contenu du cours de façon très générale. Le test ne contient aucune dimension visant des contenus de cours spécifiques et se prête bien, par conséquent, à toute la gamme des matières de cours enseignées au niveau collégial.
2. PERPE évalue le contrôle de l'apprentissage dans la mesure où celui-ci est traditionnel, c'est-à-dire fait au moyen d'examens administrés par le professeur.
3. PERPE explore minutieusement la forme de présentation du contenu (Section II: structuration, synthèses, exemples, etc...) faite par le professeur, de même que les caractéristiques de la présentation orale (Section III: débit, tonalité, capte l'attention, etc...). Le test s'applique donc primordialement aux situations d'enseignement magistral.
4. Les qualités interpersonnelles et intellectuelles du professeur sont en nombre limité et seules quelques qualités, qui à la fois discriminent bien entre les individus et ont une résonance pédagogique, ont été retenues.
5. PERPE évalue les attitudes pédagogiques propres à un enseignement centré sur l'étudiant.
6. Le test mesure quelques aspects hors cours de la relation professeur-étudiant.

Ce que PERPE n'évalue pas:

1. PERPE dans sa forme actuelle n'évalue pas d'aspects spécifi-

ques du contenu d'un cours, ni certains aspects du travail de l'étudiant (lectures, travaux) susceptibles de varier beaucoup d'un secteur à l'autre.

2. PERPE n'évalue pas la perception des étudiants quant à des formes inusitées d'évaluation telles qu'auto-évaluation ou évaluation par le groupe.

3. Dans sa forme actuelle, PERPE n'évalue pas de façon détaillée des méthodologies nouvelles de travail et d'enseignement telles que travaux d'équipe, cours vidéo, cours programmé, recherches ou stages.

4. Les qualités professionnelles ou éthiques des professeurs telles que conscience professionnelle, engagement apparent face à la profession et compétence académique, de même que leurs caractéristiques physiques ne sont pas évaluées par PERPE.

5. Les aspects disciplinaires du rôle du professeur sont minimisés et mesurés par deux dimensions seulement.

Rappelons que les principaux motifs de ces limites du contenu, en dehors de quelques réels oublis, sont l'essai et l'élimination de plusieurs dimensions pour des raisons psychométriques, et la trop grande rareté dans le milieu de certains aspects de la situation d'enseignement. Il ressort que les dimensions du test représentent adéquatement le prototype de plus courant de l'enseignement collégial actuel: un enseignement où le professeur joue un rôle prédominant, où la connaissance est surtout présentée par lui, sous forme habituellement verbale; un cours où la communication entre professeur et étudiants est coutumière et au sein de laquelle le professeur s'efforce de centrer son cours sur l'étudiant, de façon à faciliter sa démarche personnelle.

B - LA PRÉSENTATION DU CONTENU

Le contenu du test doit répondre à une double préoccupation:

a) une préoccupation d'ordre théorique, en fonction de laquelle nous devons nous assurer que la situation d'enseignement est bien représentée, en principe, par le répertoire des dimensions;

b) une préoccupation d'ordre concret, en fonction de laquelle nous devons nous assurer que chaque situation concrète d'enseignement peut être représentée, en pratique, par les résultats.

L'examen de la nature même du contenu (partie A, ci-haut) nous a permis de répondre à la première préoccupation.

L'aptitude du test à bien représenter, non pas l'enseignement en général, mais telle situation d'enseignement, dans son individualité, est fonction des qualités formelles du contenu. En effet, c'est la forme donnée au contenu qui encadre les réponses des sujets. Il faut donc s'assurer que le mode de présentation du contenu ne comporte pas des suggestions ou des influences indirectes qui biaisent les résultats.

Plus précisément, la forme de présentation de chaque dimension doit nous permettre d'assumer validement que toutes les personnes qui inscrivent leurs réponses évaluent la même chose. Cette condition est essentielle à l'interprétation des résultats, puisque ceux-ci reposent sur des moyennes de groupe. De plus, l'unification des références internes des évaluateurs est importante en fonction du fait que leur observation n'est pas immédiate; ils n'ont pas à observer et à évaluer une période de cours se déroulant actuellement sous leurs yeux. Ils doivent plutôt faire appel à leur mémoire et se représenter le cours dans son ensemble; les activités de recherche et de synthèse qui sont ainsi demandées aux sujets introduisent suffisamment d'éléments personnels pour qu'il soit indispensable de les encadrer le plus rigoureusement possible.

Les qualités formelles du contenu seront discutées dans l'ordre suivant:

1. l'opérationalité des dimensions
2. l'unidimensionalité des dimensions.

1. L'opérationalité des dimensions.

> 1.7. Chaque titre de dimension est accompagnée d'une description concrète se rapportant à des faits vécus en classe. Au moyen de cette méthode, nous unifions le contenu mental des évaluateurs.
>
> 1.8. Chacun des points des échelles d'évaluation est décrit pour aider les étudiants à juger de façon précise et objective.

Qu'il soit question du professeur (Section III et VI du test), de son enseignement (Sections I et II) ou de l'interaction professeur-étudiants (Sections IV et V), le titre d'une dimension est toujours suivi d'une définition opérationnelle, c'est-à-dire d'une description comportant des faits vécus en classe.

Exemples: 31. INTELLECTUELLEMENT STIMULANT: le professeur pousse l'étudiant à s'interroger, à réfléchir personnellement, à répondre lui-même aux questions qu'il pose, à faire des recherches personnelles. Il sait "motiver", "donner le goût" du travail et du défi à relever.

44. FRÉQUENCE DES ÉCHANGES D'OPINIONS: nous voulons évaluer ici la fréquence des interventions en classe qui dépassent en envergure les échanges du type "question-réponse". Il peut s'agir tout aussi bien de discussions entre étudiants ou avec le professeur, d'apports d'objections, d'expressions spontanées d'expériences personnelles, etc....

56. JUGEMENT NUANCÉ: ce titre réfère au professeur qui, à l'intérieur de son cours, sait introduire des nuances, peser le pour et le contre; il voit toujours les implications d'un problème, ne fait pas de généralisations hâtives ni ne tranche trop rapidement les questions.

Nous avons choisi à dessein d'illustrer, par le choix de ces trois exemples, une dimension dont même le titre est quantitatif ("fréquence des échanges d'opinion") et au moins deux dimensions visant des aspects qualitatifs du professeur, pour montrer que même des qualités aussi intangibles que "stimulation intellectuelle" ou "jugement nuancé" peuvent être opérationalisées. En définissant opérationnellement les titres des dimensions, l'interprétation personnelle que peuvent faire les répondants est réduite au minimum.

Les échelles aussi doivent être opérationnelles, en ce sens qu'elles doivent s'en tenir le plus possible aux faits, tout comme les descriptions. Il est beaucoup plus difficile d'élaborer des échelles "concrètes", que de décrire concrètement des dimensions, parce que tous les aspects qualitatifs du cours ou du professeur supposent des jugements de valeur de la part des étudiants. Comment s'y prendre pour éviter de donner prise à une trop grande subjectivité? Tout d'abord des échelles purement qualitatives, dans le genre "1. mauvais, 2. médiocre, 3. moyen, 4. bon, 5. très bon" ou "insatisfaisant, moyen, satisfaisant" furent exclues. D'autre part, la possibilité d'offrir le choix d'une simple valeur numérique fut écartée; le principe d'une échelle numérique fut conservé, mais en adjoignant une description à chaque point de l'échelle. La tâche de l'étudiant, lorsqu'il répond, consiste donc à choisir une description plutôt qu'à simplement choisir une valeur numérique située entre 1 et 5. Ce procédé empêche l'évaluateur de situer sa réponse arbitrairement ou indifféremment aux points 4 ou 5 s'il a une haute idée du professeur, ou aux points 1 ou 2 s'il en a une faible opinion.

Par exemple, lorsqu'un étudiant estime que le professeur donne peu d'informations sur l'organisation du contenu (dim. 11) il pourrait aisément situer celui-ci au point 1 sur une échelle de 1 à 5, l'évaluation subissant alors le contrecoup de l'insatisfaction éprouvée par l'étudiant. Cependant, la lecture de l'échelle descriptive contraint à la réflexion et à l'objectivité: l'étudiant doit se demander si les informations sur l'organisation du contenu sont en fait "1.très rares", 2.rares", 3.occasionnels", 4.fréquentes" ou "5.très fréquentes". Face à l'échelle, il se peut qu'il doive reconnaître que les informations sont au moins "3.occasionnelles", et que son insatisfaction est fonction de ses attentes très élevées (points 4 ou 5). En somme, la présence des descriptions oblige l'évaluateur à pondérer son jugement, de façon à ce qu'il soit le moins émotif possible.

Dans l'exemple que nous venons de fournir, il s'agit d'un comportement somme toute facile à identifier. Les informations sur l'organisation du contenu sont souvent annoncées par le professeur et elles sont distinctes du contenu. Pour un comportement de ce genre, il est plutôt aisé de construire une échelle de fréquence. Cependant, même des échelles rattachées à des aspects aussi qualitatifs que "intellectuellement stimulant (dim. 31)" ou "attitude valorisatrice (dim. 37)" sont "ramenées", si l'on peut dire, à des choix de fréquences.

Exemples: 31. INTELLECTUELLEMENT STIMULANT: le professeur pousse l'étudiant à s'interroger, à réfléchir personnellement, à répondre lui-même aux questions qu'il pose, à faire des recherches personnelles. Il sait "motiver", "donner le goût" du travail et du défi à relever.

1. TRÈS RAREMENT stimulant
2. RAREMENT stimulant
3. OCCASIONNELLEMENT stimulant
4. SOUVENT stimulant
5. TRÈS SOUVENT stimulant

37. ATTITUDE VALORISATRICE: la présence de cette attitude valorisatrice se traduit par les comportements suivants: le professeur encourage les étudiants, les félicite de leurs succès ou progrès, accepte les personnes telles qu'elles sont. A l'inverse, on ne la retrouve pas chez le professeur qui déprécie les étudiants, qui pointe, attaque ou ridiculise des personnes.

1. attitude RAREMENT valorisatrice
2. attitude OCCASIONNELLEMENT valorisatrice
3. attitude SOUVENT valorisatrice
4. attitude TRÈS SOUVENT valorisatrice
5. attitude TOUJOURS valorisatrice

2. L'unidimensionalité des dimensions.

1.9. Chaque échelle de mesure est univoque, c'est-à-dire ne mesure qu'une réalité ou un seul aspect d'une réalité complexe.

1.10. En somme, nous avons plusieurs garanties qui font en sorte que les gens évaluent bien la même réalité.

Le fait que PERPE mesure 61 aspects précis de la situation d'enseignement dépend de ce que chaque dimension ne mesure qu'une réalité. En ce qui touche les descriptions, ceci implique qu'elles sont restreintes à la seule dimension en cause: elles décrivent un seul comportement, ou elles décrivent la face négative et la face positive d'un même comportement. Si nous revenons aux deux descriptions illustrées à la fin de la section précédente, nous pouvons voir que la dim. 31 (**Intellectuellement stimulant**) décrit le professeur qui est stimulant; tandis que la dimension 37 (**Attitude valorisatrice**) décrit aussi bien celui qui ne possède pas cette attitude que celui qui la possède. Aucune règle précise n'a orienté la construction de descriptions pour faire en sorte qu'elles soient majoritairement unipolaires ou bipolaires. Le test comprend 40 dimensions unipolaires, dans le style de la dimension 31 illustrée ci-haut, et 21 dimensions bi-

polaires, dans le style de la dimension 37. Les dimensions bipolaires sont apparues en réponse à une exigence de clarté et sont utilisées, chaque fois que nécessaire, pour mieux centrer l'attention des évaluateurs sur un même objet précis d'évaluation.

Du côté des échelles, un grand soin fut apporté à ce qu'elles ne traitent que d'une seule et même réalité. Les résultats d'expérimentation permettent de contrôler ceci, puisque les dimensions non univoques provoquent de trop grandes dispersions de réponses. Ainsi, la dimension 08 (Type de contrôles) de la forme B faisait usage, dans la description et dans l'échelle, des notions de compréhension et de mémorisation dans les contrôles, selon l'hypothèse que ces deux notions pourraient constituer les deux pôles d'une même attitude. Les réponses furent assez dispersées pour suggérer l'existence de deux réalités distinctes; conséquemment, deux dimensions furent introduites dans la forme C: "Compréhension dans les contrôles" (dim. 21) et "Mémorisation dans les contrôles" (dim. 39). Les classes répondirent avec plus de cohésion et d'ailleurs la corrélation entre ces deux dimensions s'avéra assez faible (-0.39) pour justifier qu'elles aient été scindées.

Lorsque les descriptions sont bipolaires, l'échelle ne reprend qu'un seul des pôles de la description, comme en témoignent les exemples qui suivent.

Exemples: 12. APPROCHE PÉDAGOGIQUE: à l'un des pôles se situe le professeur qui ne livre presque jamais lui-même les connaissances, mais recourt à d'autres techniques pour amener l'étudiant à acquérir ces connaissances (questions en classe, travail d'équipe, travail personnel à la bibliothèque, lecture de textes polycopiés, etc...). L'autre pôle est occupé par le professeur qui présente lui-même son contenu sous forme d'exposé.

1. livre TOUJOURS lui-même des connaissances
2. livre TRÈS SOUVENT lui-même des connaissances
3. livre SOUVENT lui-même des connaissances
4. livre OCCASIONNELLEMENT lui-même des connaissances
5. livre RAREMENT lui-même des connaissances

48. PONCTUALITÉ ATTENDUE: Il s'agit ici d'évaluer dans quelle mesure le professeur se montre sensible et même strict devant la non-ponctualité des étudiants. À un pôle se situe le professeur très strict, qui pourra même refuser systématiquement l'entrée du cours dès la cloche sonnée; à l'autre, le professeur qui semble totalement indifférent face à la non-ponctualité des étudiants.

1. TRÈS PEU STRICT
2. PEU STRICT
3. PLUS ou MOINS STRICT
4. PLUTÔT STRICT
5. TRÈS STRICT

L'échelle de la dimension 12 (**Approche pédagogique**) ne joue que sur le terme "livre lui-même ses connaissances"; l'échelle de la dimension 48 (**Ponctualité attendue**) ne joue que sur le terme "strict". Les points de l'échelle se distinguent entre eux uniquement par une variation d'intensité de la réalité visée.

Dans l'ensemble, toutes les échelles sont constituées par des changements de fréquence ou d'intensité plutôt que par des changements qualitatifs[1]. Ainsi, la rigueur de formulation des échelles complète l'opérationalité des dimensions elles-mêmes. La conjonction de ces deux qualités formelles, unicité et opérationalité, rend aussi homogène que possible le contenu mental des personnes qui évaluent une même dimension, et garantit d'autant plus l'objectivité des observations.

---

[1] Six échelles font exception à l'unidimensionalité. Les extrémités de ces échelles, cependant, sont bien les deux pôles d'une même réalité. Il s'agit des dimensions 21. Syntaxe et vocabulaire (Expression verbale de plutôt pauvre à extrêmement riche), 23. Débit (de très lent à très rapide), 24. Régularité (de très irrégulier à très régulier), 28. Tonalité (de très monotone à très vivant), 43. Responsabilité face à la fréquence réduite des questions (de beaucoup plus souvent attribuable au professeur à beaucoup plus souvent attribuable à l'étudiant) et 47. Forme de discipline (de interventions très fermes à interventions très discrètes).

C - L'APPRÉCIATION DU CONTENU PAR LES PROFESSEURS

Après l'envoi des dossiers individuels de la forme B, vingt-cinq professeurs du Collège de Maisonneuve furent interrogés au moyen d'une entrevue dirigée et d'une question écrite. Suite à l'expérimentation de la forme F, 90 professeurs des six CEGEP où le test PERPE fut administré répondirent à un questionnaire-sondage.

Les questions portaient sur les impressions générales des professeurs face à PERPE, les raisons du choix des classes soumises au test, leur évaluation des résultats et leur attitude face à un éventuel service PERPE. Les réponses aux questions portant sur les impressions générales et sur l'évaluation des résultats constituent une source extérieure de jugement sur la **validité de l'instrument et de sa méthode. Il s'agit ici de la vérification auprès des usagers du test, mentionnée au chapitre II (p. 7) comme voie utile pour la validation du contenu.**

1. Les entrevues de 1968 (forme B).

1.11. Les témoignages spontanés de 52% des professeurs assurent que PERPE est valide.

1.12. L'attitude de 92% des professeurs est positive ou très positive.

1.13. Le dossier individuel offre des confirmations d'intuitions latentes ou de perceptions claires pour 61% des professeurs.

1.14. Les résultats du dossier correspondent en général aux attentes des professeurs.

1.15. Un changement de comportement est provoqué par le dossier individuel chez 75% des professeurs.

1.16. Les réponses des professeurs ne sont jamais fonction de l'excellence de leurs résultats.

La première question, non-structurée, demandait "de résumer en quelques mots ce que vous pensez du test PERPE". Les réponses spontanées de 52% des professeurs peuvent être regroupées sous le thème "bonne validité des résultats". Les résultats correspondent, selon eux, à la réalité. L'hypothèse qu'une perception positive soit reliée à de bons résultats, et qu'une perception négative corresponde à des résultats décevants, fut vérifiée au moyen d'une comparaison entre les cotes globales de satisfaction des professeurs faisant des critiques positives, et celles des professeurs faisant des critiques négatives (tests "t"). Il n'y a pas de différence significative quant à l'insatisfaction suscitée par les professeurs de l'un ou de l'autre groupe. Ainsi les témoignages spontanés de 52% des professeurs interrogés acquièrent d'autant plus de poids qu'ils ne comportent aucun lien avec l'excellence des résultats.

Sur une question structurée et soumise par écrit aux professeurs après les entrevues, ceux-ci n'expriment aucune réserve sur des aspects fondamentaux du test. En effet cette question demandait de choisir l'un des 6 points d'une échelle d'attitudes envers PERPE; 92% des professeurs se situent aux points 1 (attitude très positive: aucune ou très peu de réserves) et 2 (attitude positive, mais avec quelques réserves).

Les professeurs développent une attente vis-à-vis PERPE lorsqu'ils ont lu le répertoire des dimensions et qu'ils ont administré le questionnaire à leur classe. Au cours de l'enquête nous leur demandions (Question 5): "Le dossier individuel qui vous a été remis a été pour vous l'occasion:

1 - de nombreuses découvertes inattendues
2 - de quelques découvertes inattendues
3 - de confirmations d'intuitions latentes
4 - de confirmations de perceptions déjà claires."

La distribution des réponses est donnée au tableau 3. Il est intéressant de noter pour la validité de PERPE que 61.1% des professeurs voient dans leurs résultats une confirmation de leur auto-évaluation. Les étudiants ne sont donc pas éloignés de la "réalité" de leurs professeurs, au dire même de ces derniers. Il n'est pas négligeable non plus que 36% des

Tableau 3

Distribution des réponses à la question 5 de l'enquête PERPE (1968)

	Fréquence	%	% fusionné
1	1	2.8	38.9
2	13	36.1	
3	17	47.2	61.1
4	5	13.9	

professeurs fassent quelques découvertes inattendues; s'ils ne faisaient que des découvertes, ils ne pourraient accorder beaucoup de foi au test; s'ils n'en faisaient aucune, PERPE n'aurait pas d'intérêt pour eux. D'autre part, nous jugeons que si 47.2% des professeurs trouvent confirmation d'intuitions latentes, ils obtiennent le meilleur type d'information que PERPE peut fournir: une connaissance d'eux-mêmes déjà pressentie mais plus claire et plus certaine que celle que leur procure leur auto-évaluation. Il fut vérifié de nouveau que l'excellence des résultats n'est nullement liée au fait qu'un professeur fasse plutôt des découvertes inattendues que des confirmations d'intuitions latentes.

Devant estimer dans quelle mesure le niveau global de satisfaction exprimé par les étudiants correspondait à leurs attentes (Q.6), les professeurs pouvaient répondre sur une échelle en 5 points, allant de 1 (très au delà de mes attentes) à 5 (très en deçà de mes attentes). Quarante-cinq pour cent (45%) d'entre eux choisissent le point 3 (correspond à mes attentes), alors que le pourcentage fusionné des personnes utilisant les points 1 et 5 est de 10% seulement. Quant aux différences observées entre les 2 classes d'un même professeur (Q.15), 71% des professeurs ayant administré PERPE à 2 classes répondent au moins "en grande partie" à leurs attentes.

Ces réponses aux questions 6 et 15 montrent à nouveau que l'évaluation étudiante rejoint fort bien l'auto-évaluation des professeurs. Enfin, l'estime qu'ont les professeurs pour le profil PERPE fourni par leur classe est confirmée dans le fait que 79% d'entre eux répondent (Q. 9) qu'ils ont "pris en note une ou quelques dimensions dont le résultat était plus faible dans le but exprès de travailler à les corriger". Parallèlement, 71% répondent (Q. 16) que leurs résultats au test PERPE ont contribué à modifier à long terme certains ou plusieurs de leurs comportements comme professeurs.

2. L'enquête de 1969 (forme F).

1.17. Les professeurs de 1969 envisagent PERPE comme moyen d'évaluation objective.

1.18. Les dossiers individuels correspondent aux attentes des professeurs et servent à modifier leur comportement, autant et plus en 1969 qu'en 1968.

1.19. Soixante-seize (76%) des professeurs aimeraient utiliser de nouveau le test PERPE.

Les 196 professeurs des 6 CEGEP qui ont administré PERPE au printemps '69 reçurent en octobre de la même année un questionnaire pouvant exiger 30 à 40 minutes de leur temps. Quatre-vingt-dix (90) d'entre eux, soit 46%, y répondirent. Quelques questions nous donnent une indication de l'attitude des professeurs face à la validité de PERPE. Les questions ne sont pas les mêmes qu'en 1968, puisque le test était mieux connu dans le milieu, et que nous cherchions par le sondage à obtenir des informations plus précises que la première fois.

Ayant le choix de répondre sur une échelle allant de "1. pas du tout" à "4. beaucoup", 70% des professeurs répondent que leur décision d'administrer PERPE est "4. beaucoup influencée par la perspective d'obtenir une évaluation objective[1] par les étudiants de mon rendement pédagogique" (Q. 5). Par ailleurs, 84% des professeurs affirment avoir pris en note des dimensions précises dans le but de modifier leur comportement ou leurs attitudes (Q. 16). Ainsi, tout comme en 1968, les professeurs témoignent clairement en faveur de la validité qu'ont à leurs yeux les résultats de l'évaluation.

La correspondance entre les résultats obtenus et les attentes des professeurs se maintient, comme en 1968, à un niveau élevé. Ainsi les cotes globales de satisfaction et les différences entre les résultats de deux classes (Q. 9 et Q. 14) correspondent "à peu près" aux attentes des professeurs, dans des proportions de 61% et de 68% respectivement. Parallèlement au fait que PERPE correspond assez bien aux attentes des professeurs, les profils en stanines (Gagné, 1969, page 36), basés sur l'insatisfaction exprimée par les étudiants, contiennent "1. aucune" ou "2. très peu" de surprises dans 58% des cas (Q. 10). Bref, d'une année à l'autre, les étudiants font preuve de la même capacité à fournir aux professeurs un profil et des résultats qui sont vraiment "croyables" (Q. 9, 10 et 14), qui sont considérés comme "objectifs" (Q. 5) mais qui contiennent en même temps assez d'informations nouvelles pour que les professeurs veuillent modifier et améliorer leur comportement (Q. 16).

Ce bilan positif de l'attitude des professeurs face à PERPE se résume bien par leurs réponses à deux questions générales. D'une part, 96% d'entre eux jugent (Q. 25) que l'utilisation du test PERPE fut "1. enrichissante" ou "2. intéressante". D'autre part, 76% affirment qu'ils utiliseraient sûrement le test PERPE s'il était de nouveau offert cette année (Q. 26).

---

[1] Nous soulignons.

Ces confirmations '68 et '69 de la validité de PERPE ne ressortent pas d'une méthodologie statistique imposante, mais elles répondent aux conditions d'une bonne validation, parce que notre critère externe est parfaitement en ligne avec l'un des objectifs majeurs du test. En effet, si nous visons à procurer aux professeurs un instrument de perception sociale qui puisse adéquatement enrichir leur perception d'eux-mêmes, l'appréciation de ceux qui ont subi l'épreuve constitue une source essentielle de renseignements et de validation.

D - SYNTHÈSE

Un bref regard sur l'origine des dimensions nous a permis de constater que la construction du test a épousé d'aussi près que possible des situations réelles d'enseignement, telles que vécues par des étudiants de niveau collégial. La problématique d'une formulation consonnante au langage et aux schèmes de référence des étudiants fut résolue par l'utilisation des travaux des étudiants.

L'analyse logique des dimensions élaborées au moyen de ce procédé empirique permet de répartir celles-ci en trois zones d'égale importance: le cours (le message), le professeur (l'émetteur) et l'interaction (la communication). Le contenu du test ne porte pas sur l'étudiant (le récepteur). L'examen détaillé des sections du test permet de faire ressortir les dimensions comme échantillon valable d'un enseignement à la fois magistral quant à la provenance du savoir, mais centré sur l'élève et débordant la salle de cours quant à l'interaction professeur-étudiant.

L'opérationnalité et l'unidimensionalité des dimensions nous permettent d'assumer que les références internes des répondants sont guidées par des repères assez précis pour qu'ils évaluent la même chose et que leurs réponses soient comparables.

En définitive, le contenu même du test illustre bien la relation professeur-étudiant, envisagée du point de vue de l'étudiant, et telle que vécue actuellement dans la plupart des cours; par ailleurs, la forme du contenu est telle que les réponses des élèves d'une classe peuvent être considérées comme illustrant le même aspect d'un cours donné. Les témoignages de deux importants échantillons de professeurs démontrent concrètement que le test mesure effectivement ce qu'il entend mesurer.

Chapitre II

LA VALEUR DES ÉTUDIANTS EN TANT QU'OBSERVATEURS

Les qualités psychométriques d'un test sont habituellement circonscrites par les indices de validité et de fidélité; lorsqu'il s'agit d'une échelle d'observation, des qualités supplémentaires liées aux difficultés de la tâche des observateurs doivent aussi être envisagées.

A - LES ÉTUDIANTS COMME OBSERVATEURS PRIVILÉGIÉS

2.1. Une bonne mesure d'observation exige que l'observateur ait un échantillon représentatif des conduites à évaluer.

2.2. L'étudiant a l'avantage d'être témoin de toute la conduite d'enseignement de son professeur.

Les échelles de mesure fondées sur l'observation directe du comportement d'une personne ou d'un groupe risquent d'être inutiles lorsqu'on ne peut s'assurer que l'observateur obtient un échantillon représentatif des comportements ou situations que l'on se propose de mesurer. Un observateur venant de l'extérieur pour étudier l'interaction entre un professeur et ses étudiants, s'il assiste à un nombre restreint de cours, risque d'être témoin, sans le savoir, de comportements inhabituels. Son statut de personne extérieure l'empêche alors d'être à même d'établir une comparaison, et il ne pourra évaluer de façon juste. Ou encore la présence elle-même d'une personne extérieure risque de modifier le ou les comportements observés (Medley & Mitzel, 1963, p. 306). Plusieurs études reposant sur l'évaluation des professeurs par un observateur extérieur peuvent être critiquées à cause du caractère partiel de l'observation.

Pareil danger est absent des mesures PERPE, puisque l'étudiant, par définition, est l'observateur permanent de son professeur; il assiste à toutes les leçons du cours évalué et connaît les variations de techniques d'enseignement qui peuvent survenir selon que le cours débute, se développe ou se termine. Il connaît, non pas un simple échantillon, mais la totalité des comportements pédagogiques de son professeur. Dans l'optique des conditions d'une bonne observation, les étudiants doivent être considérés comme les observateurs privilégiés de leurs professeurs.

B - L'APTITUDE DES ÉTUDIANTS À RÉPONDRE AU QUESTIONNAIRE

2.3. Un observateur qui utilise une échelle de notation doit bien connaître les types de comportement à mesurer et comprendre à fond l'échelle de mesure.

2.4. Des critères techniques (contradiction, cohésion, etc...) de sélection des dimensions ont servi à éliminer les dimensions mal comprises des étudiants.

2.5. Un contrôle permanent sur la compréhension est effectué en écartant des compilations les étudiants qui s'abstiennent de répondre à 25% des questions.

2.6. Les professeurs évalués par leurs étudiants en 1968 et en 1969 n'ont pas mis en doute leur compréhension des questions.

L'observation pose un problème spécifique de notation, puisque la tâche exigée doit correspondre aux capacités de l'observateur. Il ne faut pas, par exemple, demander à l'observateur de juger des compétences qui lui sont étrangères, ni lui fournir des catégories de notation qui ne

lui sont pas familières. La familiarisation avec la nature même des comportements à évaluer, et la maîtrise des points de l'échelle avec laquelle se fait l'évaluation, constituent ordinairement les objectifs majeurs d'un plan de formation des observateurs. Dans le cas de PERPE, puisque les observateurs ne sont pas spécialement entraînés en vue de la notation, il est important d'établir au préalable une concordance entre le niveau de la tâche et la capacité des étudiants en tant qu'observateurs.

Nous sommes à même de constater maintenant que les exigences de rigueur qui ont présidé à la construction du test et à la sélection des dimensions ont servi non seulement à valider le contenu des dimensions qui furent conservées, mais aussi à rejeter les dimensions qui risquaient par leur ambiguité d'être mal comprises. Le fait que dans la construction de la forme A les étudiants aient eux-mêmes suggéré les aspects importants de la relation professeur-étudiant nous assure que les dimensions mesurées se situent dans des schèmes de référence qu'ils connaissent bien. Suite aux diverses expérimentations, les dimensions qui donnaient lieu à des taux de contradiction (PCD) trop élevés et à des rejets trop nombreux étaient écartées, ces critères pouvant être considérés comme des indices d'incompréhension ou de confusion chez les étudiants-observateurs. La dimension 46 de la forme A, intitulée "Digressions" fournit l'exemple d'une dimension rejetée parce qu'elle donnait des signes de l'incompréhension des étudiants: elle accusait un indice de contradiction moyennement élevé (PCD = .48) en même temps qu'une très faible cohésion intra-classe quant à l'idéal (écart-type de B = .87 en moyenne). Ces deux indices obtenus <u>simultanément</u> laissent voir que les étudiants s'entendent peu sur le niveau souhaité de digressions parce que selon toute probabilité certains les trouvent désirables et d'autres indésirables. Pareille dispersion des attentes entraîne nécessairement contradiction au sein du groupe quant à la prédominance de A sur B ou vice-versa. Nous pouvions en déduire que les étudiants comprenaient la dimension chacun selon des schèmes de référence personnels. A partir de ces constatations la dimension "Digressions" fut éliminée pour donner lieu à deux dimensions plus compréhensibles: "Digressions désirables" et "Digressions indésirables". De 0.48 à la forme A, les PCD de la

forme D sont tombés à 0.18 et 0.21; tandis que de 0.87 dans la forme A, la cohésion est passée à 0.78 et à 0.60 dans la forme D.

La dimension 19 de la forme B, appelée "Degré de dépendance face à l'attitude plus ou moins réceptive du groupe", illustre un phénomène comparable quant à l'incompréhension de nos "observateurs". Malgré son paragraphe descriptif, elle donnait lieu à des dispersions intra-classe très fortes de l'évaluation de la réalité (MSA = 1.05) et de l'évaluation de l'idéal (MSB = 1.00), aussi bien qu'un taux de contradiction très élevé (PCD = 0.48). La dispersion systématique des réponses à l'intérieur des classes montre bien que chacun comprenait la dimension comme il le pouvait et un tel état de fait nécessita que la dimension soit supprimée, ce qui fut fait.

Un autre exemple complétera l'illustration des contrôles exercés pour écarter les dimensions sujettes à l'ambiguité ou à la non-compréhension. La dimension 15 de la forme A, appelée "Réducteur de tension", occasionnait très peu de contradiction (PCD = 0.04), contrairement aux deux dimensions discutées ci-haut. Par contre, une dispersion intra-classe fort élevée sur l'évaluation de la réalité (MSA = 0.93) était accompagnée de 12% de rejets. Dans cette dimension un problème de compréhension fait en sorte que 12% des étudiants n'exécutent pas leur tâche de notation. La dimension fut abandonnée.

Conscient de la nécessité d'une vérification méticuleuse de la compréhension des étudiants face à leur tâche d'observation et de notation, d'autant plus qu'il s'agit d'observateurs non-entraînés, l'auteur a examiné chaque dimension pour y déceler des indices d'incompréhension, au moyen des PCD, des rejets ou des dispersions des mesures. Toutes celles qui paraissaient boiteuses sous ce rapport furent écartées ou transformées (1).

---

(1) L'ensemble des critères utilisés pour la sélection des dimensions sera discuté dans le texte B - 310.

Les dimensions conservées sont considérées comme correspondant au moins à la capacité moyenne de compréhension des étudiants. Dans les cas individuels d'incompréhension, les premières expérimentations ont montré que l'étudiant s'abstient de répondre. Lorsque le pourcentage d'abstentions d'un individu atteint 0.25, il est légitime de supposer qu'une raison de ces abstentions est l'incompréhension de l'étudiant. Il est plus prudent dans ce cas de mettre en doute la validité des réponses données et d'écarter complètement le sujet de la compilation des résultats. Ce contrôle est effectué de façon permanente pour chaque administration du test PERPE.

Les enquêtes menées auprès des professeurs ayant subi le test PERPE en 1968 et en 1969 n'ont laissé apparaître aucune remarque à l'effet que leurs étudiants seraient inaptes à comprendre le questionnaire ou à y répondre. Au contraire leur appréciation systématiquement positive de PERPE, en connaissance de cause puisqu'ils avaient été évalués par leurs étudiants, confirme clairement la valeur du pari qui fut fait au départ sur les capacités des étudiants comme observateurs.

C - L'OBJECTIVITÉ DES ÉTUDIANTS

> 2.7. Un instrument de mesure est qualifié d'objectif lorsqu'il n'est pas fonction des caractéristiques particulières de l'évaluateur.

L'objectivité des évaluateurs est l'une des propriétés essentielles d'une échelle de mesure, presqu'au même titre que la fidélité ou que la validité. Remmers (1963, p. 330) la définit comme la capacité que possède l'instrument de n'être pas "fonction des caractéristiques particulières de l'évaluateur". Par caractéristique particulière il faut en-

tendre ici des traits qui ne sont pas censés intervenir comme variables dans l'établissement des résultats. Dans le cas de l'échelle PERPE, les caractéristiques reliées aux étudiants sont de plusieurs ordres: la réussite académique, le sexe, l'âge, la matière étudiée, le secteur, le statut obligatoire ou non du cours évalué. Il est légitime de faire l'hypothèse que des caractéristiques d'ordre pédagogique telles que la matière, le secteur ou le statut du cours interviendront de façon peut-être significative dans les résultats, d'autant plus que dans l'optique de PERPE il s'agit là de variables sociales: la matière, le secteur, etc..., sont des variables du groupe-classe beaucoup plus que des variables individuelles.

Par contre, il est capital qu'une variable individuelle comme le succès académique ne soit pas reliée de façon trop univoque aux résultats. Si c'était le cas, le profil d'un professeur ne serait plus fonction de son comportement mais de la réussite ou de l'échec de chaque étudiant.

1. Réussite scolaire et satisfaction.

2.8.	La satisfaction générale qu'exprime un étudiant à l'égard de son professeur n'a aucun lien avec son succès dans la matière enseignée par ce professeur.
2.9.	La satisfaction exprimée par un étudiant n'a pas non plus de lien avec sa réussite moyenne dans l'ensemble de ses cours.
2.10.	Le rendement moyen de la classe dans la matière évaluée présente une corrélation de -0.42 avec la satisfaction exprimée envers le professeur.
2.11.	En somme, l'objectivité individuelle des étudiants est assurée mais un léger effet de groupe intervient dans le niveau de satisfaction d'une classe.

Tableau 4

Corrélation entre le rendement dans la matière évaluée
et la satisfaction globale (MAB) au niveau des individus

Classe	r	N	signif.*
33	0.20	47	n.s.
27	-0.06	35	n.s.
49	-0.09	34	n.s.
59	0.07	32	n.s.
75	-0.09	33	n.s.

* n.s.: non-significatif

La relation entre le rendement et la satisfaction globale de chaque étudiant fut étudiée avec les données des 88 classes de la forme B. Pour chaque étudiant, son rendement dans la matière évaluée et son rendement global étaient classés dans l'un des six niveaux: moins que 60%, de 60 à 64%, de 65 à 69%, de 70 à 74%, de 75 à 79% et 80% et plus. La mesure de satisfaction utilisée fut la MAB (satisfaction brute), calculée pour chaque étudiant sur l'ensemble des 61 dimensions. La corrélation r de Pearson entre le rendement dans la matière évaluée et la satisfaction globale fut calculée pour 5 des 88 classes; les résultats sont présentés au tableau 4.

L'absence de corrélation ou même de tendance vers une corrélation entre le rendement dans la matière et la satisfaction rend superflu de pousser les calculs au delà de 5 classes. La satisfaction générale qu'exprime un étudiant à l'égard d'un professeur n'a aucun lien avec son succès dans la matière enseignée par ce professeur. Cette absence de corrélation constitue notre plus sérieuse garantie de l'objectivité des étu-

Tableau 5

Corrélation entre le rendement moyen
et la satisfaction globale (HAB), au niveau des individus

Classe	r	N	signif.*
18	0.02	26	n.s.
60	-0.11	41	n.s.
39	-0.15	22	n.s.
24	-0.19	35	n.s.
57	-0.04	33	n.s.

* n.s.: non-significatif

diants. Il faut se demander toutefois si l'excellence générale d'un étudiant pour l'ensemble de ses cours ne fera pas de lui un étudiant plus satisfait, tandis qu'un étudiant médiocre aurait tendance à manifester plus d'insatisfaction. Cette possibilité fut vérifiée dans 5 classes par un r de Pearson entre le rendement global de chaque étudiant et son insatisfaction; les résultats sont donnés au tableau 5.

Les indices de corrélation sont de nouveau si bas qu'aucune tendance ne peut être décelée. L'objectivité des étudiants en tant qu'individus se trouve ainsi solidement confirmée, puisque ni leur rendement au cours ni leur rendement moyen n'influence leur appréciation d'un professeur.

Les constatations quant au fonctionnement individuel des étudiants peuvent-elles être transposées au niveau des groupes? Les corrélations entre le rendement moyen d'une classe (dans la matière évaluée d'une part et pour l'ensemble des cours d'autre part) et la satisfaction globale sont présentées au tableau 6. Elles portent sur l'échantillon

Tableau 6

Corrélations entre le rendement moyen des classes et la satisfaction globale (MAB)

	r	N	signif.
Moyenne de classe dans la matière et MAB	-0.42	88	.01*
Moyenne de classe en général et MAB	-0.14	88	n.s.

* .27 est significatif à .01

complet (88 classes) de la forme B. La moyenne de la classe en général n'offre pas de corrélation avec la satisfaction globale; par contre, le rendement de la classe dans la matière évaluée présente une corrélation négative avec la satisfaction exprimée à l'endroit du professeur. Plus la classe a un niveau élevé de réussite dans la matière du cours évalué, moins elle est insatisfaite. Cette observation ne porte pas directement atteinte à l'objectivité des observateurs puisque notre préoccupation ici est surtout de justifier leur objectivité individuelle. Il convient toutefois de remarquer que les observateurs sont soumis à un léger effet de groupe; lorsque la classe dans son ensemble a des notes scolaires élevées, la classe a tendance (rappelons que r = -0.42 seulement) à être plus satisfaite de son professeur.

2. Sexe et satisfaction.

2.12. La comparaison des réponses entre garçons et filles pour la forme B révèle que les filles évaluaient l'idéal et la réalité

plus haut que les garçons sur certaines dimensions.

2.13. La plupart des dimensions dénotant des différences entre garçons et filles sont disparues de la forme F.

2.14. Les classes étant presque toujours mixtes, il est probable que ces différences influencent pas les scores globaux.

L'identité sexuelle influence-t-elle beaucoup l'évaluation du professeur? A première vue la situation pédagogique suggère plutôt une hypothèse de non-différence entre l'évaluation faite par des garçons et des filles. Les indications fournies par l'application du test "t" aux MA, MB et MAB des garçons et des filles pour chaque dimension de la forme B sont présentées au tableau 7. Ce tableau permet de voir pour chaque mesure le nombre de dimensions présentant des différences significatives et le sens des différences observées entre garçons et filles.

Tableau 7

Dimensions révélant des différences significatives (test "t") au niveau de MA, MB, MAB et MAL entre garçons et filles

	N*	%	à .01		à .05		Au total		N. de dim. disparues de la forme F
			G > F	F > G	G > F	F > G	G > F	F > G	
MA	13	21	1	3	3	6	4	9	9/13
MB	28	45	2	15	2	9	4	24	12/28
MAB	11	18	3	0	7	1	10	1	6/11

* Nombre de dimensions impliquées

Des différences significatives apparaissent dans 21% des dimensions pour l'évaluation de la réalité et dans 45% des dimensions pour l'évaluation de l'idéal. Dans le cas de MA comme dans le cas de MB, les filles ont une moyenne plus élevée: elles ont des attentes plus exigeantes dans 24 sur 28 dimensions significativement différentes; elles évaluent la réalité de façon plus élevée dans 9 sur 13 dimensions. Garçons et filles diffèrent quant à la satisfaction (MAB) sur un très petit nombre de dimensions (11 sur 61 ou 18%) et 10 fois sur 11, les garçons sont plus insatisfaits que les filles. Il semble donc que nous devions conclure que l'identité sexuelle intervient réellement comme variable dans le jugement et le mode d'évaluation de nos observateurs.

Les classes étant toutefois mixtes la plupart du temps dans nos CEGEP, il est probable que les différences de scores entre garçons et filles jouent un rôle mineur dans la composition des scores globaux. L'utilisation de deux types d'observateurs est cependant légitime lorsque leurs différences de fonctionnement sont connues, et ces différences seront explorées dans le texte C - 510. Il est certain que les précisions que nous pourrons apporter aux différences d'évaluation entre garçons et filles ne pourront qu'enrichir la signification de nos résultats et notre compréhension de la psycho-pédagogie différentielle.

3. Age et satisfaction.

> 2.15. L'âge des étudiants n'a aucun lien avec leur niveau de satisfaction.

L'utilisation d'étudiants comme observateurs fait surgir la question de leur âge. Lorsqu'ils terminent le cours secondaire et entrent au CEGEP, ils ont en moyenne 16 à 17 ans. Lorsqu'ils terminent en CEGEP III ils peuvent avoir jusqu'à 20 ans. Il est plausible d'envisager que la dis-

crimination et le jugement critique exigés par la tâche seraient mieux exercés à la fin du cours qu'au début, et qu'une différence d'âge d'un à trois ans produise des différences de jugement chez des adolescents. L'hypothèse que les plus jeunes seraient plus globalement satisfaits parce que moins exigeants et moins critiques se présente spontanément à l'esprit.

La corrélation r de Pearson établie entre l'âge et la satisfaction globale (MAB) pour les 88 classes de la forme B est de 0.06 (n.s.). Il semble donc certain qu'aucun biais de jugement, soit positif soit négatif, n'intervienne chez nos observateurs en fonction de leur âge.

4. L'importance des dimensions.

2.16. Les jugements portés par les professeurs et les étudiants sur l'importance des dimensions ont des corrélations allant de 0.62 à 0.67.

2.17. Ce parallélisme du jugement des étudiants avec celui des professeurs est une indication de leur objectivité.

Les dimensions ne sont pas toutes d'égale importance et plusieurs techniques ont été utilisées depuis le début du projet PERPE pour connaître l'importance relative accordée à chacune d'elles par les professeurs et par les étudiants.

Nous trouvons une indication indirecte de l'objectivité des étudiants au moyen de l'importance qu'ils accordent aux diverses dimensions. Nous avons pu constater que leur évaluation des dimensions ressemble à celle des professeurs. Avant d'administrer la forme A, le même questionnaire fut distribué à 22 professeurs et à 180 élèves du Collège de Maisonneuve, leur

Tableau 8

Corrélations entre les cotes pondérées des professeurs de Trois-Rivières, Rimouski et Sherbrooke et les étudiants de Bois-de-Boulogne*

	TR	RK	SH
TR			
RK	.94		
SH	.85	.89	
Etudiants	.67	.62	.64

* dl 56; r significatif à .01 = .32

demandant de situer chaque dimension sur une échelle de 5 points. Les tests "t" de différence entre les moyennes d'évaluation sur chaque dimension ne sont pas significatifs dans 44 cas sur 53, c'est-à-dire pour 83% des dimensions jugées.

Lors des expérimentations C et D, les étudiants choisissaient les 5 dimensions les plus importantes, tandis que lors de l'expérimentation F, les professeurs indiquaient les 9 dimensions les plus importantes. L'une et l'autre méthode ont donné lieu à une pondération des dimensions et les corrélations entre les cotes pondérées des professeurs de 3 CEGEP en 1969 avec celles des étudiants de 1968 à Bois-de-Boulogne sont présentées au tableau 8.

Les professeurs manifestent une entente presque parfaite et plus élevée entre eux qu'avec les étudiants, mais les corrélations entre étudiants et professeurs sont toutes hautement significatives, puisque déjà

un r de Pearson de 0.32 est significatif à 0.01. Nous estimons que la capacité qu'ont les étudiants d'évaluer les éléments de la relation pédagogique de la même façon que les professeurs est une garantie de leur objectivité lorsqu'ils évaluent les professeurs.

En conclusion, les indications que nous avons de la compréhension des étudiants, de leur capacité à utiliser les échelles de notation et de leur objectivité sont suffisamment positives pour que leur rôle dans l'évaluation des professeurs soit entièrement justifié. La valeur des observateurs en tant que premiers instruments de mesure étant assurée, nous trouvons là une sérieuse garantie de la validité de PERPE.

Chapitre III

LA VALIDITE DE CONSTRUIT

3.1. Un construit est une réalité psychologique déduite à partir des comportements observables.

3.2. La validité de construit doit justifier l'hypothèse qu'un auteur fait quant à l'existence d'une variable théorique.

3.3. Le seul véritable construit du test PERPE est la notion de satisfaction.

La validité de construit d'un test entre en ligne de compte lorsqu'un auteur "se propose d'interpréter le test comme mesure d'une variable théorique" (A.P.A., 1966, p. 23). Par variable théorique il faut entendre ici un processus psychologique dont on n'a qu'une connaissance indirecte, mais dont on suppose qu'il joue un rôle déterminant dans l'obtention des scores fournis au test. (Peak, 1959, p. 287). Des exemples de tels processus, connus indirectement par l'entremise de comportements observables, sont l'habileté mentale générale ou l'agressivité[1]; ce sont des concepts, ou construits, qui dans le cadre d'une théorie psychologique du fonctionnement de l'être humain, se veulent une explication d'un ou plusieurs comportements. Le problème de la validité de construit en est un spécialement de la signification que l'on peut attribuer aux résultats, c'est-à-dire aux scores, dans le cadre d'une théorie du comportement. "Quelle est la

---

[1] Le construit principal d'un test psychologique est souvent indiqué dans le nom même donné au test.

nature et la signification du processus qui a été ainsi l'objet d'une observation ou d'une inférence? Quelle étiquette lui attribuer? Quelle est sa validité?"(Peak, 1959, p. 287).

L'importance qu'il faut donner à la validité de construit du test PERPE doit tenir compte du fait que l'inférence intervient peu dans nos analyses. Notre principale activité auteur des résultats n'est pas d'en déduire des variables hypothétiques, comme c'est le cas par exemple avec des questionnaires de personnalité à partir desquels on interprète telle réponse comme indice d'anxiété ou d'insécurité. Puisque nous cherchons primordialement à dégager en profils différentiels les dimensions de l'interaction professeur-étudiant, nous colligeons les rapports d'observation sans réellement transformer le sens que les observateurs ont mis eux-mêmes dans leurs réponses. Notre principale préoccupation de validité est forcément de nous assurer au départ que nos catégories d'observations sont bien choisies (cf. chapitre I) et qu'elles nous renseignent plus sur les professeurs que sur la subjectivité des observateurs (cf. chapitre II).

Les concepts utilisés pour désigner les scores obtenus au test PERPE sont en nombre limité: nous employons les termes de réalité, d'idéal et de satisfaction. La mesure dans laquelle il est valable d'envisager les notions de réalité et d'idéal comme nous le faisons ne feront l'objet que d'une brève récapitulation, car il s'agit là bien plus d'évidences empiriques, mises à jour par les réponses des étudiants, que de véritables construits. Le construit fondamental du test PERPE est celui du continuum satisfaction-insatisfaction, et l'examen de la validité de construit portera surtout sur ce concept. Deux thèmes seront étudiés: le rationnel du score de satisfaction et les hypothèses liées au construit satisfaction. La possibilité d'envisager la satisfaction étudiante comme indice de la compétence des professeurs complétera ce chapitre sur la validité de construit.

A - LES NOTIONS DE "RÉALITÉ" ET D'"ATTENTES"

3.4. Les scores A et B représentent explicitement la perception de la réalité du cours et les attentes des étudiants.

3.5. Considérer la moyenne d'un groupe (MA) comme représentative de la réalité est une hypothèse qui repose sur a) une définition sociale de l'objectivité, b) des indices d'accord intra-classe adéquats et c) l'entente très grande entre les paires de classes d'un même professeur.

Les scores A et B caractérisent la perception des étudiants quant à la réalité d'un cours et quant à leurs attentes en regard de ce cours. Toutefois les termes réalité et idéal n'ont pas précisément la même signification selon qu'ils réfèrent aux notations fournies par un individu ou aux notations compilées de toute une classe. Au niveau individuel les termes de réalité et d'attentes furent l'objet d'instructions explicites:

> "La question A veut rejoindre votre perception de la 'réalité' de ce qui se passe de fait en classe. La question B mesure plutôt un certain 'idéal attendu' dans la situation-cours que vous évaluez présentement." (test PERPE, forme F, p. 1)

Puisqu'il s'agit immédiatement des formulations choisies par les étudiants, les scores A et B n'ont pas à être envisagés comme des construits au sens strict du terme, lorsqu'il s'agit des scores individuels. Toutefois, l'addition des évaluations A et B des élèves d'une classe pour constituer les moyennes MA et MB est un dépassement des réponses dont nous donnons ici la justification. A notre sens cette intervention sur les scores individuels n'amoindrit pas ni ne déforme le concept de réalité, mais au contraire le

renforcit et lui donne son sens le plus riche. Nous renforcissons le sens du terme réalité en passant de A à MA parce que la réponse A d'un seul individu ne peut être envisagée que comme perception: nous ignorons quelle est son objectivité et nous ne pouvons faire aucune hypothèse sur le lien entre la réponse A d'un élève et le comportement du professeur. Par contre, lorsque plusieurs réponses individuelles vont dans le même sens, nous avons de quoi fonder une hypothèse sur le professeur comme tel: nous dirons que son "vrai" comportement se situe aux environs de la catégorie perçue par la majorité de ses étudiants comme lui étant pertinente.

Ainsi, les échelles de réponses occasionnant de trop fortes dispersions de perceptions étudiantes ont été rejetées lors de la construction du test. L'écart-type moyen de MA étant 0.87, nous pouvons dire qu'en général 66% des étudiants situent leur professeur à l'intérieur de deux des cinq intervalles de l'échelle. Si c'était là notre seule mesure d'accord "inter-juge", nous aurions à souhaiter une cohésion plus forte entre les étudiants d'une même classe. Cependant, une de nos mesures de fidélité de l'instrument nous permet de comparer entre eux les deux profils fournis par deux classes distinctes pour un même professeur. Nous avons constaté que les deux profils d'un même professeur, se ressemblaient, dans 8 cas sur 10, plus étroitement que n'importe quel couple de profils de deux professeurs différents.

Ainsi, la constance de l'accord entre les deux classes d'un même professeur nous justifie amplement d'envisager les MA comme de véritables représentations objectives du fonctionnement du professeur; il convient donc de les appeler "scores de réalité" puisque la réalité se définit comme étant ce sur quoi les gens s'entendent. Dans l'expression 'perception de la réalité', constamment utilisée par nous à propos des réponses A, il faut surtout retenir l'aspect perception lorsqu'il s'agit des individus et mettre en relief l'aspect réalité lorsqu'il est question d'une moyenne de classe.

La notion d'attentes utilisée par nous pour décrire les moyennes de scores B est encore plus facile à justifier puisque l'accord des étudiants

est plus grand ici que sur A. Par ailleurs, les scores MB représentent des rapports d'introppection que font les étudiants sur une réalité personnelle interne. Les scores A sont par leur nature même appelés à changer d'un professeur à l'autre, mais les scores B ont des chances d'être plus stables si l'on assume que les besoins pédagogiques des adolescents, d'une classe à l'autre et d'un CEGEP à l'autre, sont modelés par la même formation antérieure et par la même culture ambiante. Effectivement, les étudiants varient moins d'une classe à l'autre sur B (écart-type = 0.27) que sur A (écart-type = 0.51). Lorsque la MB d'une dimension demeure stable d'un CEGEP à l'autre, nous pouvons formuler l'idéal étudiant pour une époque donnée du réseau des CEGEP et transmettre ainsi aux professeurs et administrateurs quelques notions généralisées sur ce que désirent les étudiants.

B - LE CONSTRUIT "SATISFACTION"

3.6. Le score d'insatisfaction correspond à la différence entre les valeurs numériques des réponses aux questions A (réalité) et B (attentes).

3.7. La notion de satisfaction est fondamentale à toute théorie psychologique; elle exprime la coincidence entre un apport extérieur (réalité, expérience) à la personne et une représentation interne (désir, concept de soi) qui lui sert d'étalon.

3.8. En appelant satisfaction la superposition de A (réalité) et de B (attentes) et insatisfaction l'écart entre les deux mesures, nous adoptons une conceptualisation éprouvée.

3.9. Nous sommes justifiés d'appeler insatisfaction globale (MAB) la moyenne des écarts A - B d'un individu ou d'une classe parce que les scores MAB de dimensions ont en moyenne une corrélation de 0.62 avec la MAB globale.

3.10. Des artifices d'évaluation tels qu'effet de halo ou tendance à l'indulgence ne contaminent pas trop nos scores ni par conséquent notre notion de satisfaction.

L'explication des interprétations qu'un auteur fait des résultats de son test est spécialement importante lorsqu'il "n'accepte aucune autre mesure comme critère définitif de la qualité dont il se préoccupe" (A.P.A., 1966, p. 13). Un construit tel que celui de satisfaction étudiante n'a pas donné lieu, à notre connaissance, à des méthodes spécifiques de mesure. Remmers (1963) n'en mentionne aucune dans son chapitre sur les échelles de mesure dans les recherche pédagogiques. Ausubel (1968) n'en parle pas non plus dans son chapitre sur les facteurs de motivation dans l'apprentissage. Lorsque les professeurs sont soumis à l'observation, il s'agit presque toujours de trouver un lien entre l'activité du professeur et l'activité en classe ou les résultats des étudiants. Nous devrons donc discuter le score de satisfaction du test PERPE sans tenter de le mettre en corrélation avec d'autres scores apparentés. Nous adopterons plutôt une approche plus théorique, par laquelle nous mettrons en lumière la logique qui sous-tend la mesure.

1. Le postulat fondamental.

Le score d'insatisfaction repose sur l'établissement d'un lien entre l'évaluation de la réalité (MA) et l'évaluation des attentes face à cette réalité (MB). Il s'agit donc d'une vue synthétique des résultats et nous verrons plus loin à quel point cette méthode facilite le traitement unifié de plusieurs types de données. Le calcul de cette mesure d'insa-

tisfaction repose sur un postulat fondamental que nous pouvons formuler comme suit:

> "Toute différence entre les valeurs numériques données à une réponse A et à une réponse B doit être considérée comme indicative d'une insatisfaction chez l'étudiant."

De ce postulat découlent deux corollaires:

1) "Toute absence de différence entre les valeurs numériques données à une réponse A et à une réponse B doit être considérée comme indicative de satisfaction chez l'étudiant."
2) "Plus grande est la différence entre les valeurs numériques des réponses A et B, plus grande est l'insatisfaction; plus faible est la différence entre ces deux valeurs numériques, plus faible est l'insatisfaction" (Gagné, 1969, p. 19).

Ces corollaires nous permettent de nous prononcer avec plus de précision sur la nature des liens qui unissent les processus composants A et B.

2. Le contexte théorique.

Les références explicites au concept de satisfaction n'abondent pas dans les théories psychologiques parce que, semble-t-il, c'est un processus ou une réalité tellement fondamentale qu'elle ne peut être discutée. Une théorie de la personne humaine ne peut faire autrement que de s'articuler sur un dualisme interne-externe; en termes freudiens on pense "pulsion" ou "id" et "réalité"; en langage rogérien on discute du "moi" (self) et de l'"expérience"; en philosophie contemporaine on dira "sujet" et "projet existentiel". Ces données de base sont comme des éléments en interaction dynamique, se renforcissant ou s'opposant mutuellement. Lorsqu'ils se renforcissent, un équilibre est atteint, qu'on appellera "gratification" ou "satisfaction", "congruence de soi" ou "réalisation du sujet". Dans chaque théorie, la face externe de la personne (réalité) est confrontée à la face interne qui lui sert d'étalon: désir, représentation de soi ou conscience de soi. Ainsi, la notion d'écart entre un élément intra-psychique et une donnée extérieure est-elle toujours présente dans les formulations du dynamisme humain.

Formuler l'insatisfaction, dans nos scores PERPE, comme écart entre l'attente et la réalité répond donc aux formulations non seulement d'une mais de plusieurs écoles de pensée. Avec la notion de satisfaction nous utilisons une notion complexe, certes, mais nous croyons qu'elle coiffe admirablement les notions de réalité et d'idéal qui en sont les éléments constituants.

Si nous revenons aux corollaires mentionnées plus haut, nous constatons que les liens établis entre les mesures partielles A et B sont identiques à ceux qui peuvent exister en théorie entre l'attente et la réalité dans l'établissement d'un état de satisfaction: la coincidence des mesures est indice de satisfaction et l'insuffisance de A (réalité) par rapport à B (idéal ou attente) est indice d'insatisfaction, proportionnellement à la distance entre les deux.

3. Les applications multiples du score de satisfaction.

La moyenne des rapports A/B d'un élève, ou MAB individuelle, permet de situer chaque individu d'une classe par rapport aux autres; l'histogramme des MAB d'un groupe représente donc les variations individuelles du niveau de satisfaction. L'ensemble des rapports A/B d'une classe pour une même dimension ou MAB de dimension, permet de comparer les dimensions entre elles sous le rapport de la satisfaction procurée par le professeur aux étudiants. Toutefois, pour chaque dimension, nous calculons une seconde mesure d'insatisfaction en tenant compte du sens de la différence entre A et B. Nous appelons cette moyenne algébrique, score MAL. Chacun de ces scores, MAB individuelle, MAB de dimension ou MAL font partie de la même conceptualisation psychologique de la satisfaction.

Si les rapports A/B de chaque dimension peuvent être considérés comme des indices de satisfaction-insatisfaction, il reste que nous devons justifier la cumulation de ces mesures isolées en mesures globales, comme

c'est le cas pour la MAB individuelle et la MAB globale de classe[1]. Ces éléments observables sont-ils en corrélation suffisante pour que nous puissions parler d'un processus global de satisfaction (Nunnally, 1967, p. 86)? Avec les formes B, C et D de PERPE, la corrélation moyenne entre les MAB des dimensions et la MAB globale est 0.62. Ce niveau de consistance interne est suffisant pour assumer l'existence d'un processus général, sans être trop élevé pour que nous devions penser que les dimensions sont identiques.

4. Les risques inhérents au score de satisfaction.

Un test qui fait évaluer plusieurs caractéristiques d'une même personne à la suite l'une de l'autre court le risque de susciter un effet de halo (Remmers, 1963, p. 335), c'est-à-dire de favoriser la tendance des sujets à fixer leur niveau d'évaluation. Cet effet de halo contre lequel Remmers nous met en garde serait lié au fait que l'examen de toutes les caractéristiques d'une même personne, vues successivement, subirait l'influence d'un jugement global, positif ou négatif, porté sur elle. Dans ses suggestions pour la préparation d'échelles d'évaluation, Remmers (idem, p. 335) recommande de faire évaluer plusieurs objets (ou personnes) sur une même caractéristique pour éviter l'effet de halo. Un procédé de ce genre est impraticable dans un test comme PERPE, puisque nous pourrions au maximum demander à l'étudiant d'évaluer ses six professeurs simultanément. Avec les 61 dimensions de PERPE, le test deviendrait beaucoup trop lourd à administrer. Nous croyons parer partiellement aux dangers de l'effet de halo, au moyen de l'alternance constante imposée à l'étudiant par l'exigence de formulation successive de deux points de vue distincts à propos du même objet: réalité et attentes.

---

[1] "Il est justifié d'utiliser une notion d'ensemble... lorsque les attitudes reliées entre elles sont consistantes." (Green, 1954, p. 335).

Avec deux séries obtenues séparément nous aurions pu discuter longtemps s'il est légitime ou non d'envisager pour chaque dimension le rapport des deux mesures comme un rapport d'insatisfaction. Notre méthode a l'avantage certain de ne pas comparer deux séries de mesures à l'insu du sujet, car il doit avoir lui-même conscience du rapport qu'il établit. Cette confrontation maintenue entre attente et réalité a par ailleurs le mérite de réduire la tendance à l'indulgence, mécanisme bien connu de l'évaluation d'autrui (Remmers, 1963, p.341), étant donné que la simultanéité des deux points de vue encourage l'écart des mesures. Il peut même sembler que la forme du test suggère presque une attitude négative puisqu'elle facilite l'attribution de scores différents pour l'idéal et pour la réalité. Nous constatons que de fait l'identité des deux scores survient dans 55% des cas, assez souvent donc, pour que nous ne puissions attribuer ni aux étudiants ni aux instructions un biais systématique qui rende notre construit artificiel. Pour mesurer une attitude, il faut théoriquement pouvoir mesurer un changement d'attitude, chez un même individu à propos d'un même objet (Peak, 1963, p. 329), et l'alternance A/B à chaque dimension nous paraît faire précisément cela: mesurer un changement de point de vue quant il y a lieu, sans obliger le sujet à manifester un changement qui serait dû seulement à la forme du test.

Notre construit satisfaction, en somme, est suffisamment éprouvé théoriquement et assez clair pour que la signification de nos résultats, dimension par dimension, ne soit pas ambigue. Les dimensions sont d'autre part assez homogènes au plan de l'insatisfaction pour qu'il soit valide de parler de satisfaction globale. Le mode de présentation du test permet adéquatement d'éviter les embûches de l'effet de halo et de la tendance à l'indulgence, de sorte que notre construit n'est point trop contaminé par des artifices d'évaluation.

C - QUELQUES HYPOTHÈSES LIÉES AU CONCEPT DE SATISFACTION

3.11. Si notre notion de satisfaction est bonne, nous pouvons prévoir certaine caractéris-

tiques des résultats.

3.12. L'hypothèse que les professeurs les plus satisfaisants ont des scores MA plus élevés et non pas des MB plus basses que celles des professeurs insatisfaisants est confirmée.

3.13. L'hypothèse que les étudiants marginalement insatisfaits ont les mêmes attentes mais sous-évaluent la réalité lorsqu'on les compare à leur classe est partiellement confirmée. Ils sous-évaluent systématiquement leurs professeurs mais, en même temps, ils ont souvent des exigences plus élevées.

L'utilisation d'un concept théorique doit normalement donner lieu à des hypothèses sur la conduite des personnes qui ont un score faible ou élevé au test qui mesure ce concept. La formulation de telles hypothèses, la collecte de données pour les vérifier et l'évaluation de la capacité qu'a la théorie pour rendre compte des données recueillies sont autant d'étapes de la validation de construit (A.P.A., 1964, p. 13). La collecte de données pour vérifier une hypothèse théorique implique le recours à un critère extérieur puisqu'il s'agit d'établir des liens entre le score à l'étude et d'autres mesures qui lui seraient théoriquement reliées. Cette démarche nous est impossible, vu l'absence de mesures connues de satisfaction, mais nous pouvons prévoir l'orientation que devraient prendre certains de nos résultats, si notre formulation du construit satisfaction est bonne. Nous ferons deux prévisions, l'une portant sur la distinction entre professeurs satisfaisants et insatisfaisants, et l'autre, sur les étudiants marginalement insatisfaits.

Hypothèse 1: les professeurs satisfaisants et insatisfaisants.

Nous avons déjà défini l'insatisfaction comme l'écart éprouvé par l'étudiant entre la réalité d'un cours et son attente vis-à-vis de ce cours.

Cet écart variera selon que l'une des deux mesures s'éloigne de l'autre. Cependant, nous devons nous attendre à ce que la mesure des attentes soit relativement plus stable que la mesure de réalité, à l'intérieur d'une même population étudiante. En effet, les attentes expriment en quelque sorte l'idéal pédagogique des étudiants; il s'agit donc d'une "limite" du comportement des professeurs, par définition moins susceptible de varier que le score de réalité qui, lui, représente une description apte à refléter les différences individuelles entre les professeurs. Par ailleurs, les attentes reflètent des données intra-psychiques, par définition plus stables que des données extérieures. Il est beaucoup plus facile pour plusieurs individus de s'entendre sur un idéal que d'agir tous de la même façon ou d'évaluer la réalité tous de la même façon. En tenant à l'esprit cette stabilité de MB, nous pouvons faire l'hypothèse suivante:

> Hypothèse 1: les professeurs les plus satisfaisants se distinguent des professeurs non-satisfaisants par des scores de réalité (MA) plus élevés et plus proches des scores d'idéal (MB), eux-mêmes peu changeants, plutôt que par des scores d'idéal plus bas et moins exigeants.

Pour vérifier cette hypothèse nous avons choisi d'analyser les scores des professeurs de 2 sur 6 CEGEP de l'échantillon 1969. Les 2 CEGEP en question sont Rimouski et Sherbrooke, parce que l'utilisation des échelles de mesure, pour A et pour B, étant semblable dans chaque collège, les résultats peuvent être légitimement regroupés. Soixante (60) professeurs ont été utilisés; quinze (15) professeurs ayant des staninos 8 et 9 constituent notre groupe "satisfaisant"; vingt-neuf (29) professeurs ayant un stanine 5 forment un groupe "moyen" et seize (16) professeurs ayant un stanine 1 ou 2 constituent notre groupe "insatisfaisant". Les résultats propres à chacun de ces groupes sont présentés au tableau 9.

Il est facile de voir que la MA accuse une diminution très nette lorsque de haut en bas nous passons du groupe satisfaisant au groupe insatisfaisant: elle passe de 3.62 à 2.87. Par contre, la MB passe seulement de 3.84 à 3.76. Nous présentons au tableau 10 les valeurs "t" pour chacune

Tableau 9

Scores moyens des professeurs satisfaisants, moyens ou insatisfaisants de deux collèges* de la forme F (1969)

GROUPES	VALEURS	MA	SA	MB	SB
"S"	M	3.62	0.782	3.84	0.707
satisfaisant	σ	0.126	0.049	0.106	0.057
(stan. 8 et 9)	N	15	15	15	15
"M"	M	3.35	0.84	3.79	0.73
moyen	σ	0.147	0.050	0.121	0.060
(stan. 5)	N	29	29	29	29
"I"	M	2.87	0.931	3.76	0.770
insatisfaisant	σ	0.177	0.038	0.088	0.058
(stan. 1 et 2)	N	16	16	16	16

* Note: ces deux CEGEP ont été choisis en raison à la fois du nombre élevé de classes dans chacun et de l'utilisation similaire par les étudiants des échelles A et B.

Tableau 10

Comparaison entre les scores MA et MB de professeurs satisfaisants (S), moyens (M) et insatisfaisants (I), au moyen de tests "t".

DONNEES	COMPAR.	VALEURS "t"	dl	SIGNIF.
MA	S - M	6.05	42	.001
	S - I	13.19	29	.001
	M - I	9.76	43	.001
MB	S - M	1.23	42	n.s.
	S - I	2.95	29	.005
	M - I	2.17	43	.025

des différences. Toutes les différences d'évaluation de la réalité entre les groupes sont significatives à .001. Notre prévision des différences MA et du sens de ces différences se trouve donc pleinement confirmée. Quant à notre prévision sur la non-différence des MB, nous pouvons constater que contrairement à notre hypothèse, l'évaluation des attentes (MB) est significativement différente entre les deux groupes extrêmes: professeurs satisfaisants et professeurs insatisfaisants, de même qu'entre professeurs moyens et insatisfaisants. Toutefois la direction surprenante de la différence confirme l'ensemble de notre hypothèse, puisque les professeurs satisfaisants, plutôt que d'avoir une MB plus basse, ont une MB significativement plus élevée que celle de leurs confrères moins satisfaisants. Ils sont donc loin d'être artificiellement "satisfaisants" parce que bénéficiant d'attentes plus indulgentes, car ils sont au contraire soumis à des exigences plus grandes de la part de leurs étudiants.

Cette comparaison entre groupes extrêmes de professeurs appuie donc la signification que nous donnons au mot satisfaction: les professeurs que nous appelons satisfaisants ne sont pas ainsi classés parce que leurs étudiants ont des attentes moins élevées; au contraire ils ont un fonctionnement mieux coté, approchant de plus près un idéal qui lui-même varie peu, ou du moins, qui ne varie pas de façon à les avantager.

Hypothèse 2: les étudiants insatisfaits.

Les MAB individuelles calculées au début des analyses de résultats de chaque classe permettent de voir la distribution des élèves quant à leur niveau moyen d'insatisfaction. En utilisant 0.20 comme grandeur d'intervalle (e.g. 0.00 à 0.19, 0.20 à 0.39) nous considérons comme marginaux les étudiants séparés de la majorité de leur classe par deux intervalles ou plus.

Des cas de marginalité dans le sens d'une satisfaction exceptionnelle ne s'étant presque jamais produits à date, le terme de cas marginal est toujours utilisé pour désigner les cas d'insatisfaction déviante par

rapport au groupe. Les cas marginaux, une fois repérés, sont exclus des calculs de la classe parce qu'ils fausseraient les moyennes, mais peuvent nous être utiles pour vérifier le fonctionnement de notre construit de satisfaction.

L'insatisfaction vis-à-vis de soi-même est une dimension de la personnalité dont chacun fait l'expérience personnelle. Rogers en fait l'un des noyaux des troubles de personnalité parce qu'elle engendre, dit-il, un état de conflit interne et une perception déformée de la réalité (Rogers, 1962, p. 232). Nous pouvons nous attendre à ce que l'insatisfaction en regard d'autrui produise un effet semblable de perception déformée et entraîne un dérèglement systématique de l'évaluation d'autrui. Notre hypothèse sur le processus d'évaluation des étudiants que nous nommons marginaux se formule comme suit:

> Hypothèse 2: les étudiants marginalement insatisfaits sont caractérisés par une sous-estimation constante du fonctionnement de leur professeur par rapport à leur groupe de classe, plutôt que par des attentes plus élevées.

Pour vérifier cette hypothèse, quatorze (14) cas marginaux du CEGEP de Sherbrooke (1969) et dix-huit (18) du CEGEP de Trois-Rivières ont été comparés à leurs classes respectives, pour MA et MB. Les résultats présentés au tableau 11 montrent que les MA des sujets marginaux, en plus d'être très significativement différentes, sont toujours inférieures à celles de leurs classes. L'insatisfaction marquée de ces cas marginaux a donc bien la signification prévue: ils dévaluent systématiquement le professeur. La portion de notre hypothèse portant sur la non-différence entre les MB n'est pas vérifiée puisque ces cas dénotent aussi des MB différentes de celles de leurs confrères: 21 cas sur 32 ont une MB personnelle plus élevée que leur classe. Il peut paraître surprenant que la MB varie significativement entre les étudiants insatisfaits et les autres, puisque la MB est habituellement stable et que la différence entre les moyennes des deux groupes est seulement 0.23 dans un CEGEP et 0.13 dans l'autre.

Tableau 11

Différences entre les MA et MB de cas marginaux
et celles des classes correspondantes (1969)

CEGEP	Données	Cas marginaux	Classes correspondantes	"t"	signif.
SH	MA	2.43	3.28	5.96	.001
	SMA	0.46	0.22		
	N	14	14		
TR	MA	2.44	3.29	9.31	.001
	SMA	0.35	0.13		
	N	18	18		
SH	MB	4.03	3.80	2.44	.02
	SMB	0.32	0.09		
	N	14	14		
TR	MB	3.86	3.73	1.73	.05
	SMB	0.30	0.07		
	N	18	18		

En fait, de si petites différences sont significatives justement parce que les MB sont stables et que les écarts-types des MB sont 0.09 et 0.07 pour les deux groupes de classes impliqués. Les variations de MB sont donc beaucoup moins importantes que les variations de MA entre marginaux et groupes de classe et il demeure justifié de dire que le concept d'insatisfaction conserve avec les cas marginaux le sens habituel de déficience de la réalité en regard de l'idéal.

Disons pour conclure que notre conceptualisation de l'insatisfaction, ainsi que l'approche technique adoptée pour la mesurer, nous ont

permis de répartir aussi bien les professeurs que les étudiants en sous-groupes distincts sur le continuum satisfaction-insatisfaction. Les groupes obtenus sont caractérisés par des différences de mesures qui demeurent cohérentes avec notre postulat de base et ses corollaires, appuyant ainsi fermement la validité des notions utilisées pour caractériser les données PERPE.

D - LA NOTION DE COMPÉTENCE

3.14. La notion de compétence peut être considérée comme étant parallèle à celle de satisfaction si l'on croit que la satisfaction étudiante est un objectif valable de l'enseignement.

3.15. L'Association américaine des recherches en éducation considère la satisfaction étudiante non seulement comme un critère valable de compétence professorale mais aussi comme l'un des moins passagers.

3.16. Le fait que notre mesure MAB reflète plutôt des différences de comportement que des différences d'idéal étudiant justifie qu'on la considère comme critère de compétence.

3.17. Le rôle du professeur étant largement de disposer l'étudiant à l'apprentissage, l'attitude positive de l'étudiant est déjà un gage de la réussite du professeur.

3.18. Le test PERPE conservera sa pleine valeur s'il est utilisé pour développer la compétence, par le perfectionnement qu'il favorise chez le professeur. Il peut difficilement favoriser le perfectionnement personnel s'il est utilisé comme instrument de sélection ou de promotion.

Nos données sont surtout caractérisées par les notions de réalité, d'idéal et de satisfaction. Un autre concept peut-il avoir sa place dans l'interprétation des résultats? S'il paraît assuré que nos scores nous renseignent de façon immédiate sur la satisfaction étudiante face à la situation-cours, pouvons-nous joindre au concept de satisfaction celui de compétence professorale, et dire que plus une classe est satisfaite plus un professeur est compétent?

1. La satisfaction comme un des critères fondamentaux de la compétence du professeur.

Pour définir clairement ce qu'est la compétence, il faut la relier à un de ses effets, dont on dira alors qu'il sert de critère. Le choix d'un critère de compétence est lui-même déterminé par des valeurs sociales ou éducatives. "La compétence professorale signifie habituellement l'influence d'un professeur sur l'accomplissement d'une valeur quelconque" (Gage, 1963, p. 116). Le choix d'une valeur éducative découle de la théorie éducative à laquelle on croit ou il dépend tout simplement de préférences personnelles. Il est facile d'imaginer que des professeurs ne veuillent pas tenir compte de la satisfaction de leurs étudiants et qu'ils préfèrent qu'on les juge selon d'autres critères, comme leur préparation au cours ou la réussite de leurs élèves. Tout dépend des objectifs que chaque professeur se donne, face à son travail d'enseignant.

Les critères de l'efficacité d'un professeur peuvent être plus ou moins immédiats, c'est-à-dire plus ou moins proches de l'actuelle situation d'enseignement. La note scolaire de l'étudiant est l'exemple d'un critère immédiat, qui offre l'avantage d'être facile à mesurer. Cependant, plus les critères de compétence professorale sont immédiats, moins ils sont fondamentaux, parce que les objectifs réels de l'éducation sont à long terme; par exemple, une formation de l'esprit critique, une formation de la personne, etc.... Ainsi, l'un des critères les plus riches en signification serait "le succès et les réalisations d'un étudiant dans la vie"; mais ce critère est tellement éloigné de la situation d'enseignement qu'il n'est presque pas mesurable. Ce sont les difficultés inhérentes à la mesure des critères fondamentaux qui font en sorte que les recherches en éducation utilisent surtout des critères immédiats.

Afin d'encourager les chercheurs à affronter les difficultés liées aux critères "éloignés", l'Association américaine pour la recherche en éducation, par son comité d'étude sur les critères de compétence professorale, a publié en 1952 une liste de critères, hiérarchisés selon leur "éloignement" de la situation concrète, c'est-à-dire selon qu'ils sont plus ou moins fondamentaux par rapport à la compétence du professeur. Nous en donnons ici la traduction:

"<u>Critère ultime</u>.

Effet du professeur sur:
- le succès et les réalisations de l'étudiant dans la vie
- la réussite académique ultérieure de l'étudiant
- l'atteinte d'objectifs éducatifs actuels par l'étudiant
- la satisfaction de l'étudiant à l'égard du professeur
- la satisfaction des parents à l'égard du professeur
- la satisfaction des administrateurs à l'égard du professeur.

Les "valeurs" ou attitudes évaluatives du professeur.
La connaissance du professeur en psychologie éducative et en santé mentale.
L'adaptation émotive et sociale du professeur.
La connaissance du professeur en méthodes de construction de programmes.

La connaissance qu'a le professeur de la matière enseignée.
L'intérêt du professeur pour la matière enseignée.
Les notes du professeur pour les cours d'entraînement.
Les notes du professeur dans ses cours d'éducation.
L'intelligence du professeur."

(American Educational Research Association, 1952, pp. 243 - 244, dans Gage, 1963, p. 117)

Deux remarques sont à faire à propos de cette liste; la satisfaction à l'égard du professeur est incontestablement reconnue par ce comité d'experts comme critère valable de compétence puisqu'elle revient trois fois parmi les 6 premiers critères; en second lieu, la satisfaction étudiante est quatrième dans cette hiérarchie, ce qui la situe parmi les critères considérés comme étant les moins passagers et les plus difficiles à évaluer.

Il serait donc justifié d'envisager notre mesure de la satisfaction étudiante comme critère de compétence professorale. Cette mesure est d'autant plus apte à refléter des différences de compétence entre les individus, qu'elle varie en fonction des évaluations de la réalité, comme nous l'avons montré plus haut, et non pas en fonction des attentes des étudiants.

2. La satisfaction comme facteur de motivation pédagogique.

L'orientation moderne de la pédagogie ne manque pas par ailleurs d'accentuer l'importance de la satisfaction étudiante. Si l'on songe à la part qui est faite à la psychologie et à la psycho-pédagogie dans la formation des maîtres, on voit bien que la compétence professorale ne peut pas être réduite à l'expérience dans l'enseignement ou à la maîtrise de la matière enseignée. La connaissance des enfants, de leurs besoins affectifs autant que de leur capacité cognitive sont essentiels à la transmission efficace de la connaissance. Les écrits du psychologue Carl Rogers ont grandement contribué à centrer l'attention des enseignants sur l'étudiant comme personne plutôt que sur les notes scolaires ou sur la discipline. Ce cou-

rant de pensée, en diminuant l'importance de l'enseignement centré sur le professeur en faveur d'un enseignement centré sur l'élève, encourage une préoccupation généralisée pour les méthodes actives. Il devient alors essentiel de développer chez l'étudiant "l'intérêt pour ce qu'il apprend, en même temps qu'un ensemble d'attitudes et de valeurs appropriées envers l'activité intellectuelle en général" (Bruner, 1960, p. 73), parce qu'il est le principal agent de son propre apprentissage. Ce n'est plus le professeur qui apprend quelque chose à l'étudiant, c'est l'étudiant qui "s'apprend" des choses et le rôle du professeur devient celui d'une médiation entre l'étudiant et la connaissance. La fonction du professeur n'en est pas moins essentielle, parce que si l'intérêt pour la matière est le meilleur stimulus à apprendre, le professeur est le principal agent du développement de l'intérêt. Il serait exagéré de prétendre que la satisfaction à l'égard du professeur est la condition essentielle et unique pour que l'étudiant effectue un apprentissage. Un étudiant insatisfait de son professeur peut apprendre malgré tout s'il considère ce cours comme vital pour sa future carrière, s'il est déterminé à maintenir une moyenne élevée ou si la matière le passionne depuis longtemps. On ne peut nier par contre que les facteurs de motivation sont souvent ce qui fait le plus défaut chez les étudiants, et que des attitudes négatives peuvent même inhiber tout à fait les processus cognitifs. L'attitude positive ou la satisfaction envers le professeur ne cause pas l'apprentissage mais le facilite et l'accélère parce qu'elle dispose l'étudiant favorablement envers les objectifs et le travail proposé par le professeur. "Les variables de motivation et d'attitude ne sont pas directement impliquées dans les processus cognitifs. Ils rendent le processus plus énergique et plus rapide durant l'apprentissage en augmentant l'effort, l'attention et la disponibilité immédiate à apprendre". (Ausubel, 1968, p. 369).

3. Le risque inhérent à l'utilisation administrative de PERPE.

En somme, il paraît presqu'impossible de ne pas reconnaître l'étroite parenté qui unit la satisfaction étudiante et la compétence professorale. Nous croyons opportun, toutefois, de faire une mise en garde contre une utilisation "exclusivement administrative" du test.PERPE. La sélection

et la promotion des professeurs exigent le recours à des critères aussi objectifs que possible d'évaluation et l'on pourrait être tenté de faire intervenir les résultats du test PERPE dans ces processus. Un tel usage n'est pas contraire aux possibilités du test, mais à notre sens il provoquerait des conséquences néfastes. Pouvant prévoir les conséquences de leur jugement sur le statut de leurs professeurs, les étudiants pourraient être tentés de se concerter pour "passer sentence" sur leur professeur. L'objectivité que nous leur connaissons maintenant risquerait trop de disparaître, d'autant plus qu'en utilisant PERPE pour des fins de promotion, il faudra probablement faire passer le test dans toutes les classes vers une même époque, saturant ainsi les étudiants par une succession d'évaluations face auxquelles ils perdraient toute application. Une utilisation libre par les professeurs répartirait plus également la tâche exigée des étudiants. Un second danger réside du côté de l'attitude des professeurs face à une signification administrative des résultats. Leur attitude actuelle est extrêmement accueillante et réceptive, même chez ceux pour qui les résultats sont pénibles. Il ne fait pas de doute que le caractère tout à fait personnel du dossier crée un climat favorable à la prise de conscience des points faibles et au désir d'amélioration.

Ainsi, même si PERPE fournit des indices de satisfaction étudiante qui reflètent la compétence des professeurs, l'utilisation des informations ne doit pas nécessairement en être une de sélection ou de promotion. Au contraire, il nous paraît être un instrument propice au développement de la compétence dans la mesure où il ne sert pas à la sélection, parce que le désir de perfectionnement est impossible lorsque domine la crainte des résultats.

Chapitre IV

LA VALIDITÉ DE PRÉDICTION
ET LA VALIDATION CONCOMMITANTE

A - LA VALIDITÉ PRÉDICTIVE

> 4.1. La prédiction de la compétence des professeurs est la seule que nous pourrions envisager éventuellement.

La validité concommitante et la validité prédictive d'un test reposent toutes deux sur une comparaison entre les scores fournis au test et une variable extérieure, considérée comme fournissant une mesure directe de la caractéristique étudiée (A.P.A., 1966, p. 13).

La validité prédictive repose sur une comparaison différée entre les résultats au test et un critère externe, ce critère pouvant être les résultats fournis par un autre test, des mesures "naturelles" telles que réussite scolaire ou, pour l'industrie, le rendement dans la production. L'aspect fondamental de ce type de comparaison est de confronter des mesures obtenues à des époques différentes, afin de vérifier si une première évaluation permet de prédire le fonctionnement ultérieur d'un sujet. La validité prédictive à établir à propos de PERPE pourrait difficilement porter sur l'évaluation de la réalité, des attentes ou de la satisfaction étudiante parce que cette réalité et ces traits sont fortement liés à des circonstances toujours particulières: tel cours et telle classe. Des mesures aussi "situationnelles" que les nôtres se prêtent donc mal à une étude prédictive. Nous ne pouvons comparer les attentes (MB) ou la satisfaction (MAB) de telle classe à leurs attentes et satisfaction d'un moment ultérieur puisqu'alors un autre cours et un autre professeur seront en cause et que les mesures auront changé en fonction de la situation. Par ailleurs, les classes sont beaucoup trop mobiles au CEGEP pour que de telles prédictions aient une quelconque utilité.

S'il est inutile de vouloir prédire quelque chose à propos des classes d'étudiants, nous pourrions malgré tout tenter d'utiliser les mesures d'insatisfaction comme prédicteurs de la compétence future des professeurs. Pour mener à bien ce type de validation, deux conditions sont nécessaires.

1. Première condition d'une validation prédictive.

Il faudrait que nous puissions utiliser un autre critère, lui-même valide, de compétence professorale. Cette exigence pose un premier obstacle puisque de tels critères, comme nous l'avons dit plus haut, sont difficiles à élaborer. Cette mesure extérieure de compétence une fois trouvée, il nous faudrait connaître l'ampleur de la corrélation qu'elle peut avoir avec la mesure PERPE de compétence, en l'occurence le score global d'insatisfaction. Il serait évidemment essentiel que les deux mesures soient fortement reliées.

2. Deuxième condition d'une validation prédictive.

Puisque la validité prédictive suppose une comparaison différée entre deux types de mesure, il conviendrait de vérifier si la compétence telle qu'établie selon le critère externe demeure la même après un an ou deux d'intervalle. La valeur d'un score d'insatisfaction comme indice prédictif de compétence serait indiscutable si les mêmes personnes obtenaient, après un certain laps de temps, des scores comparables ou améliorés.

Cette étude n'est pas faite à l'heure actuelle, bien qu'elle enrichirait notre connaissance de la signification des résultats. Si nous envisageons l'état de la pédagogie au Québec, nous croyons qu'une telle recherche serait à tout le moins périlleuse, parce qu'il est difficile d'assumer que les attitudes et les techniques pédagogiques sont elles mêmes stables par nature. Pour faire une recherche prédictive en dépit des difficultés techniques, il faut être au moins justifié de croire que le trait mesuré un jour sera le même le lendemain. Or la majorité des professeurs

dans les collèges du Québec sont encore très jeunes, et il ne fait de doute pour personne que les pressions étudiantes, à elles seules, sont suffisantes pour les porter à transformer leur approche pédagogique; d'autre part, les professeurs se stimulent entre eux à expérimenter le plus possible de nouvelles formes d'enseignement.

Ce contexte naturel de changement, ajouté aux difficultés mentionnées ci-haut, rend doublement impraticable, pour le moment, toute recherche prédictive. Face à ce bilan somme toute négatif de données expérimentales quant à la valeur prédictive de PERPE, rappelons toutefois que la prédiction de la compétence ne constitue pas l'un des objectifs majeurs du test. Dans son manuel de directives aux constructeurs d'instruments psychométriques, l'American Psychological Association note: "A moins toutefois que le but... d'un test soit spécifiquement de prédire ou de se substituer à un critère, la corrélation avec un critère ne constitue pas une évaluation utile de ce test"[1](A.P.A., 1966, p. 13).

B - LA VALIDITÉ CONCOMMITANTE

4.2. La validité concommitante repose sur la comparaison simultanée entre nos résultats et l'évaluation selon un critère externe, lui-même valide.

4.3. Nous ne connaissons pas de mesures validées, des attentes et de la satisfaction étudiante pouvant servir de critère externe.

---

[1] Notre traduction.

Dans le cas de la validité concommitante, où il s'agit d'une comparaison simultanée avec un critère extérieur, nous pouvons en principe étudier trois caractéristiques de PERPE: la mesure de réalité, la mesure des attentes des étudiants et la mesure de la satisfaction étudiante. Une validation simultanée de la réalité (score MA) pourrait utiliser l'appréciation par des juges extérieures des mêmes dimensions que celles de PERPE quant au fonctionnement du professeur. Une validation de PERPE comme mesure des attentes étudiantes pourrait se faire, en principe, au moyen d'une comparaison avec d'autres mesures de leurs attentes, par exemple des questionnaires de mesure d'accord ou de désaccord avec des formulations parallèles des dimensions. La satisfaction peut être validée au moyen d'autres mesures de satisfaction, soit au niveau de chaque dimension, soit au niveau global. Ces mesures extérieures devraient être des mesures elles aussi de perceptions interpersonnelles, si l'on veut qu'une comparaison soit légitime. Ces validations n'ont pas été effectuées à date, pour deux raisons surtout. En premier lieu, le temps et les ressources nous ont manqué pour entraîner des observateurs qui auraient pu effectuer une évaluation parallèle du fonctionnement des professeurs. En second lieu, il n'existe pas, à notre connaissance, de mesures reconnues, elles-mêmes validées, des attentes et de la satisfaction étudiante.

Au cours de l'une des pré-expérimentations, nous avons pu toutefois nous procurer une mesure externe, mais "artisanale" de la satisfaction étudiante. Cette mesure nous permet de vérifier la cohérence des étudiants sur deux mesures différentes de leur satisfaction.

1. Le rang moyen comme validation concommitante de la satisfaction étudiante.

> 4.4. Le rang moyen d'un cours parmi les autres cours suivis constitue une mesure externe de la satisfaction étudiante.

4.5. La MAB globale et le rang moyen des professeurs ont une corrélation de 0.83.

4.6. Nous avons dans l'ensemble une confirmation indirecte de la validité concommitante du test PERPE.

Les formes C et D de PERPE furent accompagnées d'une question demandant aux étudiants de situer leur appréciation globale de leur professeur: "En termes de satisfaction générale, quel RANG occupe le cours que vous venez d'évaluer parmi tous les cours que vous suivez cette année? (Le rang 1 correspond au cours dont vous êtes le(la) plus satisfait(e) et ainsi de suite pour les autres cours)". Un rang moyen fut obtenu pour chaque professeur à partir des rangs assignés par chacun des élèves.

Ce score global de "satisfaction générale" constitue la seule mesure extérieure que nous puissions comparer avec les scores PERPE (MAB et MAL) d'insatisfaction. Il comporte deux avantages et un défaut. En effet, il est avantageux que cette mesure provienne de la même source que les scores PERPE, parce que la comparaison des mesures est dans ce cas réellement fondée; si nous avions, par exemple, une évaluation faite par les chefs de département de la satisfaction suscitée par les professeurs auprès de leurs étudiants, nous pourrions nous attendre à des jugements différents de celui rendu par les étudiants, sans pour autant mettre en doute la validité de PERPE. De plus, le rang global a l'avantage d'être une mesure de perception sociale, de sorte que nous comparons deux mesures de même type. Par contre, il présente l'inconvénient d'être global; il ne peut être défini par des composantes et par ce fait, il a plus de chances d'être soumis à un effet de halo: un petit nombre de dimensions positivement appréciées peuvent créer chez l'étudiant un jugement d'ensemble très positif malgré une forte proportion de dimensions négativement perçues; le même effet peut se produire pour une perception d'ensemble négative, malgré de nombreux bons points chez le professeur. Tout dépend du point de vue selon lequel un étudiant se place lorsqu'il tente de déterminer le

rang d'un professeur parmi les 6 qui composent son horaire. Finalement, la possibilité d'un effet de halo est d'autant plus grande que cette question sur le rang global fut placée à la fin du test PERPE.

Pour l'ensemble des 88 professeurs de la forme B, la corrélation de rang (Spearman) entre les MAB globales et les rangs moyens est .883. Manifestement, l'approche quelque peu différente des deux mesures ne change pas grand chose aux résultats d'un professeur et l'attitude de la clause est cohérente d'une technique à l'autre.

La mesure du rang moyen nous fournit donc un indice fort élevé de validité concommitante. Evidemment, malgré sa corrélation avec la MAB globale, le rang moyen ne saurait lui être substitué puisqu'il n'a pas l'avantage de provenir de composantes multiples et de permettre une analyse détaillée des points faibles et des points forts.

CONCLUSION

Au terme de cet examen de la validité du test PERPE, sommes-nous en droit de dire que l'instrument mesure effectivement ce qu'il est censé mesurer? Le test répond-t-il aux objectifs qui lui ont été assignés?

L'objectif le plus évident et le plus fondamental du test consiste en ceci: mesurer la relation professeur-étudiant. L'étude de la validité de contenu nous a permis de constater que la relation professeur-étudiant est adéquatement représentée par les 61 dimensions du test. Cette relation, telle que circonscrite par le répertoire des dimensions, comprend l'administration de la matière et l'évaluation de l'apprentissage par le professeur, les techniques pédagogiques de ce dernier et ses modes de transmission orale, certains aspects de sa personnalité, ainsi que les interactions d'ordre socio-émotif et d'ordre cognitif. Il s'agit dans l'ensemble des aspects propres à refléter l'actualité la plus courante de l'enseignement, plutôt que des aspects applicables à certaines matières ou à certains professeurs seulement. Dans la perspective de ce premier objectif: "mesurer la relation professeur-étudiant", l'étude de la validité de contenu nous a aussi permis de conclure que l'aspect **mesure** est adéquatement favorisé par l'opérationalité et l'unidimensionalité des dimensions. Finalement, les professeurs consultés ont largement confirmé la capacité du contenu à rendre compte de leur situation pédagogique personnelle.

L'objectif PERPE voulant que le test repose réellement, non pas sur des préjugés, mais sur les perceptions étudiantes est atteint, si l'on considère que ni la réussite scolaire personnelle ni l'âge ne contaminent les jugements portés par les étudiants.

L'étude de la validité de construit nous a permis de constater que la mesure de l'insatisfaction étudiante ne fut pas une vaine entreprise. Au contraire, la classification des scores associés à ce concept aboutit à des sous-groupes cohérents et unifiés de professeurs et d'étudiants. Au delà de cette vérification, la caractérisation constante des professeurs dit "satisfaisants" par des scores plus élevés de réalité nous a conduit à établir un lien entre compétence professorale et satisfaction étudiante. Ce rapprochement est fonction d'une optique théorique qui situe la satisfaction étudiante dans la gamme des facteurs importants de motivation pédagogique.

Finalement, l'exploration des conditions préalables des validations prédictive et concommitante nous a conduit à envisager celles-ci comme des réalisations ultérieures seulement des recherches PERPE. Toutefois, nous avons pu constater que des jugements globaux, portés par les étudiants sur leurs professeurs, confirment les scores globaux élaborés à partir des 61 mesures détaillées du questionnaire PERPE.

Au terme de cet examen critique de la validité de PERPE, les jugements positifs que nous avons pu porter sur la valeur de l'instrument nous justifient clairement d'y voir un précieux et efficace outil du perfectionnement des professeurs, un outil apte à enrichir leur perception des besoins des étudiants et leur connaissance d'eux-mêmes.

Références

American Psychological Association (A.P.A.), 1966, Standards for Educational and Psychological Tests and Manuals. Washington D.C.: A.P.A.

Ausubel, D.P., 1968, Educational Psychology, A Cognitive View. New York: Holt.

Bruner, J.S., 1960, The Process of Education. New York: Random House.

Cannell, C.F. & Kahn, R.L., 1959, L'interview comme méthode de collecte, dans Les Méthodes de recherche dans les sciences sociales, Vol. II, pp. 385 - 436 (éd. par L. Festinger et D. Katz). Paris: P.U.F.

Gage, N.L., 1963, Paradigms for Research on Teaching, dans Handbook of Research on Teaching, pp. 94 - 141 (éd. par N.L. Gage). Chicago: Rand McNally & Co.

Gagné, F., 1969, Texte A - 100: Qu'est-ce que le projet PERPE?, document polycopié no. 1169 - 02. Montréal: Projet PERPE.

Green, B.F., 1954, Attitude Measurement, dans Handbook of Social Psychology, pp. 335 - 369 (éd. par N.L. Gage). Reading, Mass.: Addison-Wesley Publ. Co.

Medley, D.M. & Mitzel, H.E., 1963, Measuring Classroom Behavior by Systematic Observation, dans Handbook of Research on Teaching, pp. 247-328 (éd. par N.L. Gage). Chicago: Rand McNally Co.

Nunnaly, J.C., 1967, Psychometric Theory. New York: McGraw Hill.

Peak, H., 1959, Problèmes d'observation objective, dans Les Méthodes de recherche dans les sciences sociales, Vol. I, pp. 285 - 349 (éd. par L. Festinger & D. Katz). Paris: P.U.F.

Remmers, H.H., 1963, Rating Methods in Research on teaching, dans Handbook of Research on Teaching, pp. 329 - 378 (éd. par N.L. Gage). Chicago: Rand McNally Co.

Rogers, C. & Kinget, G.M., 1962, Psychothérapie et relations humaines, Vol. I. Louvain: Publications universitaires.

Ryans, D.G., 1963, Assessment of teacher Behavior and Instruction, dans Review of Educational Research, pp. 415 - 439, Vol. XXXIII, no. 4.

Solomon, D., & al., 1964, Dimensions of Teacher Behavior, dans Journal of Experimental Education, pp. 23 - 40, Vol. 33, No. 1.